

BMJ Open

Words do matter: a systematic review on how different terminology for the same condition influences treatment preferences

Journal:	<i>BMJ Open</i>
Manuscript ID	bmjopen-2016-014129
Article Type:	Research
Date Submitted by the Author:	03-Sep-2016
Complete List of Authors:	Nickel, Brooke; University of Sydney, School of Public Health Barratt, Alexandra; University of Sydney, School of Public Health Copp, Tessa; University of Sydney School of Public Health, Moynihan, Ray; Bond University , Centre for Research in Evidence-Based Practice McCaffery, Kirsten; The University of Sydney, Screening and Test Evaluation Program (STEP), School of Public Health; The University of Sydney, Centre for Medical Psychology & Evidence-based Decision-making (CeMPED)
Primary Subject Heading:	Communication
Secondary Subject Heading:	Medical management
Keywords:	terminology, disease labels, treatment preferences, treatment decision making, overdiagnosis, overtreatment

SCHOLARONE™
Manuscripts

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Words do matter: a systematic review on how different terminology for the same condition influences treatment preferences

Brooke Nickel *PhD candidate*¹, Alexandra Barratt *epidemiologist*¹, Tessa Copp *PhD candidate*¹, Ray Moynihan *research fellow*^{1,2}, Kirsten McCaffery *health psychologist and professorial research fellow*¹

¹Wiser Healthcare, Sydney School of Public Health, The University of Sydney, Sydney, New South Wales, Australia

²Centre for Research in Evidence-Based Practice, Bond University, QLD 4229, Australia

Correspondence to: Ms Brooke Nickel; Wiser Healthcare, Sydney School of Public Health, The University of Sydney, NSW 2006, Australia. (email: brooke.nickel@sydney.edu.au)

ABSTRACT

Objectives: Changing terminology for low-risk, screen detected conditions has now been recommended by several expert groups in order to prevent overdiagnosis and reduce the associated harms of overtreatment. However, the effect of terminology on patients' preferences for treatment options is not well understood. This review aims to synthesise existing studies on terminology and its impact on treatment decision making.

Design: Systematic review.

Methods: Studies were included that compared two or more terminologies to describe the same condition and measured the effect on treatment preferences and/or choices. Studies were identified via database searches from inception to February 2016, and from reference lists. Two authors evaluated the eligibility of studies with verification from the study team, extracted and cross-checked data, and assessed the risk of bias of included studies.

Results: Of the 1142 titles identified, 6 studies met the inclusion criteria. Five studies were quantitative and one was qualitative. Five of the studies were of high quality. Studies covered a diverse range of conditions: ductal carcinoma in situ (3), gastroesophageal reflux disease (1), conjunctivitis (1), and a bony fracture (1). The terminologies compared in each study varied based on the condition assessed. Generally, when a more medicalised term was used to describe the condition it resulted in a shift in preference towards more invasive treatments, and/or higher ratings of anxiety and perceived severity of the condition.

Conclusions: Different terminology given for the same condition influenced treatment preferences and psychological outcomes in a consistent pattern in these studies. Changing the terminology may be one strategy to reduce patient preferences for aggressive treatment responses to low-risk conditions.

Trial Registration: CRD42016035643.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

STRENGTHS AND LIMITATIONS OF THIS STUDY:

- This is the first systematic review to synthesise the evidence on how different terminology given for the same condition impacts patients’ treatment preferences
- Only a small number of studies have examined this research question and were included in the review
- Due to the variability of terms and outcomes assessed, authors were unable to conduct a meta-analysis and pool the effects of the data
- All studies included were hypothetical, therefore patients facing a real diagnosis may respond differently

INTRODUCTION

Medical encounters can be challenging and confronting for patients, especially when they are faced with a treatment decision. Clinical communication and language is an important aspect of a medical encounter as it influences patients' understanding of their diagnosis and treatment options^{1,2}.

Decisions about treatments may be influenced by various factors including the medical terminology clinicians use to diagnose and describe conditions to patients. Describing a patients' circumstances using more medicalised terminology may lead to greater preference for more invasive treatments, whether or not the extent of the treatments are warranted.

Overdiagnosis of many medical conditions and its associated overtreatment is now widely accepted^{3,4}, and can have serious implications for patients, healthcare systems and society^{5,6}. Numerous approaches are beginning to be proposed to help combat overdiagnosis and overtreatment, including various communication strategies⁷. Changing the terminology for medical conditions may be one communication strategy to mitigate the effect of overdiagnosis and overtreatment as it has the potential to influence the effect of diagnosis and labelling of a condition and influence patients' decision making about treatment. It may encourage both patients and clinicians to more carefully consider conservative treatment options.

In particular, cancer terminology is one area where use of different terminology may greatly influence treatment decision making. The term 'cancer' is understandably frightening for people to hear and can influence their thought and action⁸, but it is now well understood that a range of disorders which include indolent to fast-growing lesions are labelled as cancer⁹. Lesions with low malignant potential are common (such as low-risk DCIS, papillary thyroid cancer and low-grade prostate cancer), and with the advent and increasing use of various screening technologies these indolent lesions and their precursors are now frequently clinically identified. This identification can lead to the condition being overdiagnosed and in turn overtreated.

Changing the terminology for these low-risk, screen detected conditions has now been recommended by several expert groups – including a National Institutes of Health state of the science conference panel and a National Cancer Institute working group – in order to prevent overdiagnosis and encourage more consideration of less invasive treatment options such as active surveillance⁹⁻¹¹. However, the effect of terminology on patients’ willingness to accept more conservative treatment options is not well understood. This systematic review aims to synthesise existing studies on terminology for medical conditions and its impact on treatment decision making, and associated psychosocial outcomes.

METHODS

Protocol and registration

The review’s protocol is registered with PROSPERO (an international prospective register of systematic reviews), registration number: CRD42016035643.

Review Question

How do different terminologies given to the same condition influence treatment decision making?

Search Strategy

A comprehensive list of search terms was developed (Appendix 1) with consultation from an information specialist and a search of relevant databases (Medline, Pre-Medline, Embase, PyscINFO, Cinhal, and PubMed) was conducted from inception to February 2016. The returned search results were screened by title and abstract independently by two researchers (BN and TC) for irrelevant articles, reviews papers, editorials or commentaries, and duplicates. An eligibility checklist was developed (Appendix 2) to guide the selection of appropriate studies. Decisions regarding inclusion and exclusion of studies was then made independently by two researchers (BN and TC) and disagreements discussed. Any further disagreement or uncertainty was discussed and verified by

two additional researchers (AB and KM). A hand search of reference lists of included studies as well as papers recommended through personal communication were also examined for relevant studies.

Inclusion and Exclusion Criteria

Studies were included in the review if they empirically measured treatment decision making as an anticipated or experienced outcome of being given or told one terminology versus another or others for a specific condition. Studies were excluded if they were reviews, editorials or commentaries, or if they assessed participants less than 18 years of age making a decision for themselves (Appendix 2).

Quality Assessment and Data Extraction

All studies that met the inclusion criteria were appraised for study quality by two authors (BN and TC) independently using two separate tools; one for quantitative studies and one for the qualitative study. The quantitative studies were appraised using a modified version of the Cochrane Collaboration's tool for assessing risk of bias which was developed by study authors (Table 1). As the quantitative studies included in this review were not clinical trials, study authors adapted the Cochrane Collaboration's tool for assessing risk of bias to appraise the studies for items relating to study design, study setting, study validity, and analysis. Study validity included allocation of participants to different terminology, blinding of participants to the study hypotheses, and reporting and measurement bias. The qualitative study was appraised using criteria adapted from the Consolidated criteria for reporting qualitative studies (COREQ) framework and results appear in Appendix 3¹².

Two categories of study quality were identified by study authors according to each study's methodological characteristics. In high-quality studies (lower risk of bias) the majority of criteria were fulfilled and done well, while in low-quality studies (higher risk of bias) the majority of criteria were not done or done poorly. Results of the risk of bias assessment for quantitative studies are shown in Table 3.

Data from the final studies included in the review were extracted independently into a standardised template by two authors (BN and TC). Any discrepancies were discussed and resolved by the entire research team. For studies where the data was not reported or was unclear, authors were contacted and responded for confirmation of details of the data. Authors were also asked to confirm the extracted data to minimise any reporting bias. All authors responded. Results from the studies were synthesised in a narrative form, as the heterogeneity of the studies and their respective outcome measures did not support pooling of results¹³.

RESULTS

Initial search results identified 1142 papers. After removal of duplicates and screening by title and abstract, 19 studies underwent full-text reviews. Six studies that reported the impact of different terminology for the same condition on treatment decision making were identified (Figure 1). Three studies reported on ductal carcinoma in situ (DCIS)¹⁴⁻¹⁶, two on common childhood conditions (gastroesophageal reflux disease (GERD), conjunctivitis)^{17 18}, and one on a bony fracture¹⁹. Five studies reported quantitative findings^{14 15 17-19} and one study reported qualitative findings¹⁶. The key characteristics of these studies and their methods are summarised in Table 2. All studies were hypothetical and involved various samples of community members who were not currently and/or previously diagnosed with the condition assessed. Of the five quantitative studies, four involved a randomised experimental design; with two using a paired sample design (cross-over) and two using an independent sample design (2x2 factorial). The qualitative study was linked to one of the quantitative studies¹⁵, as women from the study were asked if they would be willing to participate in an additional qualitative interview for further investigation of the topic.

Overall the majority of studies were of higher quality with a lower risk of bias, however one study did have a high risk of bias as the quality of reporting was low¹⁹ (Table 3). Higher quality studies had defined study populations and settings, had low selection bias and described the data collection, methods and analysis appropriately. The qualitative study included in the review was rigorous and

had appropriate methods to reach its research objectives. Researchers and the target population were defined, the sampling strategy was explained, data collection methods and data analysis were appropriate and well documented.

Since the strength of the review is the diversity of included studies, results of the studies are summarised individually by study quality category, with data on treatment outcomes summarised in Table 4. Following this is a narrative synthesis of the effect of terminology (more medicalised terms vs less or non-medicalised terms) on treatment preferences (invasive treatment preference vs non-invasive treatment preference). Importantly for each study we identify the classifications of treatment preferences and terminology and provide justification where applicable (Appendix 4).

Results from individual higher quality studies

McCaffery et al. 2015 (Ductal carcinoma in situ)

McCaffery and colleagues' study on the impact of DCIS terminology on treatment preference (immediate treatment vs watchful waiting) and women's level of concern found no significant differences in treatment preference between arm A (women who were given the term 'abnormal cells' first and then were given the term 'pre-invasive cancer cells') and arm B (women who were given the term 'pre-invasive cancer cells' first and then were given the term 'abnormal cells') with 33% and 40% of women respectively favouring treatment, $p=0.23$. However, 18% of women in arm A who were initially given the 'abnormal cells' terminology changed their preference to treatment when the terminology was switched to 'pre-invasive breast cancer cells' while only 6% changed to watchful waiting ($p=0.008$). In contrast, there were no significant changes in treatment preference in arm B when the terminology was switched the other way (9% vs 8% changed their stated preference).

Similarly this study found that there was no significant difference between arms with regards to level of concern with 49% and 44% of women indicating they would be extremely concerned $p=0.600$.

However, when the alternative term was used, women in arm A (who were initially given the ‘abnormal cells’ term and then were given the ‘pre-invasive breast cancer cells’ term) were significantly more likely to report increased concern than women in arm B (67% vs 52%, $p=0.001$).

Findings from this study were also supported by the included qualitative study conducted with a subset of women¹⁶.

Nickel et al. 2015 (Ductal carcinoma in situ)

This qualitative study investigated in-depth how different proposed terminologies to describe DCIS affected women’s treatment preferences and psychological outcomes by conducting semi-structured interviews with women of varying education, cancer screening experience and with no history of a DCIS diagnosis. Findings demonstrated that overall women preferred a diagnosis of DCIS to be communicated using terminology that did not include the term cancer, as women generally exhibited stronger negative reactions when the cancer term was used to describe DCIS compared to when a non-cancer term was used. Although concern seemed to be high overall, women displayed a high level of interest in watchful waiting when it was described to them in a hypothetical scenario as a safe and effective option, and told that they could proceed to treatment in the future if necessary.

Omer et al. 2013 (Ductal carcinoma in situ)

This study also examined the impact of DCIS terminology on women’s treatment preferences by comparing 3 identical scenarios, with the only difference being the term used to described DCIS (‘non-invasive cancer’, ‘breast lesion’ and ‘abnormal cells’). Treatment options presented were surgery, medication, active surveillance. All participants saw all scenarios, with the order of scenarios varied across participants. When DCIS was described as a non-invasive cancer 53% of participants preferred non-surgical options, whereas 66% chose non-surgical options when it was described as breast lesion and 69% chose non-surgical options when it was described as abnormal cells ($p<.001$). Although women with a previous history of cancer (other than breast cancer) and

women with high socioeconomic status more frequently chose surgery in univariate analyses, high numeracy was the only independent predictor of preference for surgical treatment in multivariate logistic regression models for all three terms: cancer (OR 2.11, 1.34-3.34 CI, $p=0.001$), lesion (OR 1.96, 1.20-3.19, $p=0.001$), abnormal cells (OR 1.63, 1.01-2.67, $p=0.048$).

Scherer et al. 2013 (Gastroesophageal reflux disease)

This study on the influence of the term GERD (versus no term or label given) on parents preferences for medication for their infant found that parents who received the term GERD in the scenario were more interested in medication than parents who did not receive that term, $F(1,165) = 6.95$, $p<.01$. To assess parental interest in antibiotics, study authors combined three highly inter-correlated questions involving parent's interest in medication: Will you give your infant this medicine? Do you think your infant needs the medicine your doctor offered? Do you think that the medicine will help your infant get better?

The study also found a significant interaction between the term GERD and report of medicine ineffectiveness, $F(1,165) = 4.52$, $p<.05$ as parents who received a GERD diagnosis were interested in medicating the infant, even if they were told that the medications are likely ineffective. By contrast parents not given a diagnosis were interested in medication only when they were not given information about medication effectiveness, thus able to assume that the medications are effective.

All parents in the study were also asked (using a 0-5 Likert scale with labelled end-points for each question) whether they were worried about their infant's health (mean=2.28, SD=1.30), thought the condition was somewhat serious (mean=2.12, SD=1.19), and were relatively unlikely to describe their infant as being sick (mean=1.87, SD=1.45). These answers were not influenced however by the presence or absence of the GERD term (all $p>.12$).

Scherer et al. 2015 (Conjunctivitis)

A similar study design conducted by Scherer and colleagues did not find an initial difference between the term 'pink-eye' and 'eye-infection' on parents' preference to medicate their infant. However, when symptoms were referred to as an 'eye infection', information about antibiotic ineffectiveness significantly reduced interest in using medication $F(1, 62) = 14.67, p < .001$. By contrast, when parents were told that the symptoms were 'pink eye', interest in antibiotics was not reduced by information about antibiotic ineffectiveness $F(1, 74) = 0.93, p = .33$.

Study authors also measured (using a 0-5 Likert scale with labelled end-points) parents' perceptions about contagiousness and belief that their child could attend child care and found that parents who received the 'pink-eye' term thought that the symptoms were significantly more contagious (mean=4.04, SD=1.30) than parents who received the 'eye-infection' term (mean=3.29, SD=1.63), $F(1, 137) = 11.21, p = .001$. Parents who received the 'pink-eye' term were also less likely to believe that their child would be allowed to go to child care (mean=0.35, SD=0.90) than parents who received the 'eye-infection' term (mean=0.93, SD=1.13), $F(1, 141) = 9.70, p = .002$.

Individual results from the lower quality study

Azam et al. 2010 (Bony Fracture)

This study found that patients' treatment expectations and perceptions of bony injuries differ based on the terminology used to describe the injury. Patients expected more invasive treatments (operation or cast) when a more medicalised term was used to describe the injury compared to a less invasive treatment (sling or heals on own), with 58% of patients expecting invasive treatments for the term 'broken bone', 42% for 'fracture' and 28% for 'greenstick fracture'. In contrast, patients given a less medicalised term had a lower preference for invasive treatments, with 26% of patients choosing an invasive treatment for 'hairline fracture', and 19% for 'crack in the bone'. This study also found that patients perceived the injury to be more severe when a more medicalised term was used

to describe the injury (average mean severity score out of 10; broken bone=6.64, greenstick fracture=5.28, fracture=4.95, hairline fracture=3.58, crack in the bone=3.28).

Synthesis of results

Table 5 summarises preferences for more invasive treatment option by type of terminology across the studies. Generally, there was a clear pattern in the same direction showing that more medicalised terminology resulted in a greater preference or interest in more invasive treatments, whether this be surgical treatment for DCIS, interest in medication or antibiotic use, and operation or cast for a bony injury, although these differences did not always reach statistical significance. In contrast, when participants were given a less medicalised or non-medicalised term a higher proportion of participants chose a non-invasive treatment option such as active surveillance for the management of DCIS, had reduced interest in potentially ineffective medication or antibiotic use, and sling or heals on own for a bony injury. Studies varied as some found a significant effect of terminology on treatment decision making while others found a significant interactions effect or within person effect and psychological outcome effects (eg. perceived severity, level of concern).

DISCUSSION

This is the first systematic review which synthesizes the evidence on how different terminology given for the same condition impacts treatment preferences. Overall this review found that when a more medicalised term is used to describe a condition, people have stronger preferences for more invasive treatment options. This finding demonstrates that different terminology used to describe the same condition can influence patient's treatment preferences.

The terminology given during a diagnosis to patients is important, as when a healthy individual becomes a patient they can immediately become more vulnerable to the words their clinicians use²⁰. Using words that generate fear or anxiety cause patients to have more difficulty making informed decisions and becoming an active participant in their care²¹. It has also been shown that the use of

interpretive terminology (eg. including the words positive or negative, or using a metaphor)^{22 23}, the terminology used to describe a treatment choice²⁴, and describing a condition with plain language terminology as compared to jargon²⁵ can have an influence on medical decision making. Together with findings from this review these studies show that language is a powerful tool that has the potential to influence patients' thoughts and actions.

Our findings are also in line with the results of other research on the effect of labelling a condition in relation to the social implications it may have for the individual, rather than its effect on medical decision making. Research has found that disease labels can result in various emotional, cognitive and physical consequences^{26 27}. Most notably, studies which have examined the effect of hypertension labelling have found that giving the label of hypertension to individuals (compared to giving no label) resulted in increased self-reported illness and absenteeism from work, and a significant increase in blood pressure^{26 28 29}. In the conjunctivitis study by Scherer and colleagues included in this review, the term 'pink-eye' was perceived as being more contagious, and parents were less likely to believe their child could go to childcare, compared with parents who received the condition described as an 'eye-infection'.

Using disease labels may also reduce a patient's sense of self-control, and therefore increase perceived severity and uptake of medications. The more medical the label the less control a person may feel over the situation, increasing their perceived severity about the condition and creating a perception that more invasive medical interventions are warranted. This was shown in both the GERD and conjunctivitis studies where there was a high level of interest in medication when a label was given, even when parents' were told that the medication was ineffective^{17 18}.

Limitations

This study was limited by the small number of included studies in the review, which also included two studies that were conducted by members of the review team. However, in light of recent and

ongoing evidence of overtreatment across a number of conditions and suggestions from leading global medical bodies to change the terminology of low-risk conditions^{10 11}, this review addressed an important and timely under-researched question.

Due to the variability of the terms and outcomes assessed in the included studies, authors were unable to conduct a meta-analysis and pool the effects of the data, and therefore a definite synthesis of results of all studies was not possible.

We found it was difficult in some studies to characterise precisely why some terms elicited stronger preferences for treatment. For example, it seemed clear that pre-invasive breast cancer cells was a more medicalised term than abnormal cells. In contrast, it was not clear to us what it was exactly about the term “pink eye” that elicited stronger parental preferences for antibiotic treatment when compared with the term “eye infection”. The author team therefore made explicit judgements about which terms were deemed more medicalised and which were not, as well as what treatments were considered invasive and what were considered non-invasive. These decisions were guided from the aims and outcomes of the studies, and followed categories used in the original studies, with the exception of the Azam paper where authors were guided by the severity ratings given to the terms (see Appendix 4). We note the Azam paper was of lower quality which made interpretations about the terms used more difficult. Nonetheless, it was very clear that in all of the studies the use of different terminology for the same condition tends to elicit different responses to treatment preferences as well as psychosocial outcomes. Further research could usefully explore more fully what characterises terms that elicit stronger preferences for more invasive treatments.

Importantly, all studies included in this review were hypothetical and were not presented to patients randomised to receive a diagnosis and treatment preferences were assessed instead of actual treatment decision making. Patients facing real treatment decisions may respond differently to those in the studies, however it is likely that in real life situations these effects may be even more pronounced. It is likely that patients would be more anxious than participants in current studies,

therefore more susceptible to judgement biases that result from more medicalised terminology or labels.

Conclusion

Against a backdrop of recent evidence of overdiagnosis and overtreatment across a number of conditions^{5 30-32} and the potential physical and psychological impacts this may have for the patient, understanding how different terminology given for the same medical condition may influence patients’ treatment preferences is important. This review demonstrates that the terminology used to describe a condition consistently influences patient preferences for treatments and related outcomes. Although further evidence to strengthen findings are needed from clinical populations, this review supports the calls for changing the terminology of conditions where the risk of progression is low. Changing the terminology in low-risk conditions or conditions with indolent clinical course may be a potential communication strategy to help shift assumptions that immediate invasive treatments are always needed, allow for better shared decision making between clinicians and patients, and the consideration of more conservative treatment options.

Acknowledgements: The authors would like to thank Philippa Bourke for her help with the literature search and Dr Kevin McGeechan for his comments on the statistical aspects of the paper.

Contributions: All authors included in the paper fulfilled the criteria for authorship. BN contributed to study design, data collection, analysis, interpretation, drafting and revising the manuscript. AB contributed to study design, analysis, interpretation, and revising the manuscript. TC contributed to data collection, analysis, interpretation, and revising the manuscript. RM contributed to interpretation and revising the manuscript. KM contributed to study design, analysis, interpretation and revising the manuscript. All authors approved the final version of the manuscript and all authors are guarantors.

Funding: This paper was written with support from the Wiser Healthcare CRE grant from the National Health and Medical Research Council (NHMRC) #1104136. BN was supported by the Sydney Catalyst Research Scholar Award. KM was supported by an NHMRC fellowship (1029241). The funders had no role in the design or conduct of the study; in the collection, analysis, and interpretation of the data; or in the preparation or approval of the manuscript.

Competing interests: All authors have completed the ICMJE uniform disclosure form at www.icmje.org/coi_disclosure.pdf and declare: no support from any organisation for the submitted work; no financial relationships with any organisations that might have an interest in the submitted work in the previous three years; AB and RM are co-chairs of the steering group for the *Preventing Overdiagnosis* scientific conferences; no other relationships or activities that could appear to have influenced the submitted work.

Data sharing: No additional data available.

REFERENCES

1. Schofield PE, Butow PN, Thompson JF, et al. Psychological responses of patients receiving a diagnosis of cancer. *Annals of oncology : official journal of the European Society for Medical Oncology / ESMO* 2003;**14**(1):48-56

2. Arora NK. Interacting with cancer patients: the significance of physicians' communication behavior. *Social science & medicine* 2003;**57**(5):791-806

3. Hoffman JR, Cooper RJ. Overdiagnosis of disease: a modern epidemic. *Archives of internal medicine* 2012;**172**(15):1123-4 doi: 10.1001/archinternmed.2012.3319[published Online First: Epub Date]].

4. Welch HG, Schwartz LM, Woloshin S. *Over-diagnosed: making people sick in the pursuit of health*. Boston: Beacon Press, 2011.

5. Moynihan R, Doust J, Henry D. Preventing overdiagnosis: how to stop harming the healthy. *Bmj* 2012;**344**:e3502 doi: 10.1136/bmj.e3502[published Online First: Epub Date]].

6. Heath I. Overdiagnosis: when good intentions meet vested interests--an essay by Iona Heath. *Bmj* 2013;**347**:f6361 doi: 10.1136/bmj.f6361[published Online First: Epub Date]].

7. McCaffery KJ, Jansen J, Scherer LD, et al. Walking the tightrope: communicating overdiagnosis in modern healthcare. *Bmj* 2016;**352**:i348 doi: 10.1136/bmj.i348[published Online First: Epub Date]].

8. Dunn BK, Srivastava S, Kramer BS. The word "cancer": how language can corrupt thought. *Bmj* 2013;**347**:f5328 doi: 10.1136/bmj.f5328[published Online First: Epub Date]].

9. Esserman LJ, Thompson IM, Reid B, et al. Addressing overdiagnosis and overtreatment in cancer: a prescription for change. *The Lancet Oncology* 2014;**15**(6):e234-42 doi: 10.1016/S1470-2045(13)70598-9[published Online First: Epub Date]].

10. Allegra CJ, Aberle DR, Ganschow P, et al. National Institutes of Health State-of-the-Science Conference statement: Diagnosis and Management of Ductal Carcinoma In Situ September 22-24, 2009. *Journal of the National Cancer Institute* 2010;**102**(3):161-9 doi: 10.1093/jnci/djp485[published Online First: Epub Date]].

11. Esserman LJ, Thompson IM, Jr., Reid B. Overdiagnosis and overtreatment in cancer: an opportunity for improvement. *Jama* 2013;**310**(8):797-8 doi: 10.1001/jama.2013.108415[published Online First: Epub Date]].

12. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. *International journal for quality in health care : journal of the International Society for Quality in Health Care / ISQua* 2007;**19**(6):349-57 doi: 10.1093/intqhc/mzm042[published Online First: Epub Date]].

13. Schünemann HJ, Oxman AD, Higgins JPT, et al. *Presenting results and 'Summary of findings' tables — results without meta-analyses*, 2011.

14. Omer ZB, Hwang ES, Esserman LJ, et al. Impact of ductal carcinoma in situ terminology on patient treatment preferences. *JAMA internal medicine* 2013;**173**(19):1830-1 doi: 10.1001/jamainternmed.2013.8405[published Online First: Epub Date]].

15. McCaffery K, Nickel B, Moynihan R, et al. How different terminology for ductal carcinoma in situ impacts women's concern and treatment preferences: a randomised comparison within a national community survey. *BMJ open* 2015;**5**(11):e008094 doi: 10.1136/bmjopen-2015-008094[published Online First: Epub Date]].

16. Nickel B, Barratt A, Hersch J, et al. How different terminology for ductal carcinoma in situ (DCIS) impacts women's concern and management preferences: A qualitative study. *Breast* 2015;**24**(5):673-9 doi: 10.1016/j.breast.2015.08.004[published Online First: Epub Date]].

17. Scherer LD, Finan C, Simancek D, et al. Effect of "Pink Eye" Label on Parents' Intent to Use Antibiotics and Perceived Contagiousness. *Clinical pediatrics* 2015 doi: 10.1177/0009922815601983[published Online First: Epub Date]].

18. Scherer LD, Zikmund-Fisher BJ, Fagerlin A, et al. Influence of "GERD" label on parents' decision to medicate infants. *Pediatrics* 2013;**131**(5):839-45 doi: 10.1542/peds.2012-3070[published Online First: Epub Date]].
19. Azam N, Harrison M. Patients' perspectives on injuries. *Emergency medicine journal : EMJ* 2011;**28**(7):601-3 doi: 10.1136/emj.2009.082032[published Online First: Epub Date]].
20. Bedell SE, Graboyes TB, Bedell E, et al. Words that harm, words that heal. *Archives of internal medicine* 2004;**164**(13):1365-8 doi: 10.1001/archinte.164.13.1365[published Online First: Epub Date]].
21. Cassell EJ, Leon AC, Kaufman SG. Preliminary evidence of impaired thinking in sick patients. *Annals of internal medicine* 2001;**134**(12):1120-3
22. Zikmund-Fisher BJ, Fagerlin A, Keeton K, et al. Does labeling prenatal screening test results as negative or positive affect a woman's responses? *American journal of obstetrics and gynecology* 2007;**197**(5):528 e1-6 doi: 10.1016/j.ajog.2007.03.076[published Online First: Epub Date]].
23. Scherer AM, Scherer LD, Fagerlin A. Getting ahead of illness: using metaphors to influence medical decision making. *Medical decision making : an international journal of the Society for Medical Decision Making* 2015;**35**(1):37-45 doi: 10.1177/0272989X14522547[published Online First: Epub Date]].
24. Rot I, Ogah I, Wassersug RJ. The language of prostate cancer treatments and implications for informed decision making by patients. *European journal of cancer care* 2012;**21**(6):766-75 doi: 10.1111/j.1365-2354.2012.01359.x[published Online First: Epub Date]].
25. Pieterse AH, Jager NA, Smets EM, et al. Lay understanding of common medical terminology in oncology. *Psycho-oncology* 2013;**22**(5):1186-91 doi: 10.1002/pon.3096[published Online First: Epub Date]].
26. Pickering TG. Now we are sick: labeling and hypertension. *Journal of clinical hypertension* 2006;**8**(1):57-60
27. Hoyt CL, Burnette JL, Auster-Gussman L. "Obesity is a disease": examining the self-regulatory impact of this public-health message. *Psychological science* 2014;**25**(4):997-1002 doi: 10.1177/0956797613516981[published Online First: Epub Date]].
28. Haynes RB, Sackett DL, Taylor DW, et al. Increased absenteeism from work after detection and labeling of hypertensive patients. *The New England journal of medicine* 1978;**299**(14):741-4 doi: 10.1056/NEJM197810052991403[published Online First: Epub Date]].
29. Rostrup M, Kjeldsen SE, Eide IK. Awareness of hypertension increases blood pressure and sympathetic responses to cold pressor test. *American journal of hypertension* 1990;**3**(12 Pt 1):912-7
30. Independent UK Panel on Breast Cancer Screening. The benefits and harms of breast cancer screening: an independent review. *Lancet* 2012;**380**:1778-86
31. Moyer VA, on behalf of the U.S. Preventive Services Task Force. Screening for prostate cancer: U.S. Preventive Services Task Force recommendation statement. *Annals of internal medicine* 2012;**157**(2):120-34 doi: 10.7326/0003-4819-157-2-201207170-00459[published Online First: Epub Date]].
32. Brito JP, Morris JC, Montori VM. Thyroid cancer: zealous imaging has increased detection and treatment of low risk tumours. *Bmj* 2013;**347**:f4706 doi: 10.1136/bmj.f4706[published Online First: Epub Date]].

Table 1. Criteria for assessment of study quality for quantitative studies

	Information Extracted
Study design	Independent-sample or paired-sample design If independent: whether groups were randomised?
Study setting	Community sample, clinics, hospital, other
Selection bias	Independent sample: Were participants allocated to each group randomly? Were samples similar in terms of important characteristics? How was randomisation done? Was randomisation done with concealment allocation? Was the study sample described? Paired-sample: Were terms randomised? How was randomisation done? Was randomisation done with concealment allocation? Was the study sample described?
Performance bias	Were participants kept blind to the study hypothesis?
Attrition bias	Was attrition or exclusions (lost to follow-up) reported?
Reporting bias	Examination of selective reporting. Did they study have a protocol?
Measurement bias	Exposure variable – describe the term used Outcome measures – how were outcomes measured? Were they validated or referenced? Confounders – reporting of additional measures and/or demographics
Analysis	Were appropriate statistical tests used to analyse data and report results?

Table 2. Characteristics of included studies

Study	Year	Country	Disease Focus	Study Design	Study Sample & Setting	Study Aims	Terms Manipulated	Treatment Outcomes Measured	Other Outcomes Measured
McCaffery et al.	2015	Australia	DCIS	Randomised experimental design (cross-over design)	269 health women from a community sample	To examine whether the use of terminology including the term cancer to describe DCIS increased hypothetical level of concern and treatment preferences	'abnormal cells' vs 'pre-invasive breast cancer cells'	Treatment preferences: treatment vs watchful waiting, measured on a 5-point Likert scale (definitely prefer treatment, probably prefer treatment, prefer to do nothing, probably prefer watchful waiting, definitely prefer watchful waiting)	Level of concern measured on a 5-point Likert scale (Extremely concerned-Not concerned at all)
McKee et al.*	2015	Australia	DCIS	Semi-structured qualitative interviews	26 women from a community sample	To understand how different proposed terminologies for DCIS affect women's perceived concern and management preferences	'abnormal cells' vs 'pre-invasive breast cancer cells' 'ductal carcinoma in situ' 'ductal intraepithelial neoplasia' 'indolent lesions of epithelial origin'	Women's qualitative responses to terminologies with and without the cancer term on level of concern and management preferences	
Mermer et al.	2013	USA	DCIS	Randomised experimental design (cross-over design)	394 healthy women with no history of breast cancer from a hospital patient registry	To examine how women respond to terminology for DCIS without the cancer term	'non-invasive breast cancer', 'breast lesion', 'abnormal cells'	Treatment preferences (choice between surgery, medication, active surveillance)	
Scherer et al.	2013	USA	GERD	Randomised experimental design (2x2 factorial design) and above	175 parents aged 18 years presenting at a primary care pediatric clinic	To determine if the disease label GERD influences parents' perceived need to medicate an infant	'GERD' vs no label 'this problem'	Parents' interest in giving their infant medication (3 measures using 6-point Likert scale: No, definitely not-Yes, definitely)	Perception of illness severity (3 measures using 5-point Likert scale: worry = Not at all worried-Very worried, serious = Not at all serious-Very serious, sick = Strongly disagree-Agree) Appreciation of medication offered (1 measure using 5-point Likert scale: No, definitely not-Yes, definitely)

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49

Scherer et al.	2015	USA	Conjunctivitis	Randomised experimental design (2x2 factorial design)	159 parents aged 18 years presenting at a primary care pediatric clinic	To determine whether the 'pink eye' term would influence parents' beliefs about the condition and their interest in using antibiotics	'pink eye' vs 'eye infection'	Parents' decision to give their child antibiotics measured on a 6-point Likert scale (No, definitely not-Yes, definitely)	Targeted beliefs about contagiousness measured on a 5-point Likert scale (Not at all contagious-Very contagious) Parents ability to send their child to day care measured on a 5-point Likert scale (Not at all likely-very likely)
Sam et al.	2010	UK	Bony fracture	Cross sectional survey	100 adult patients presenting at an emergency department	To assess the way different terms used to describe a fracture affect the understanding a patients has of it including the perceived severity of the injury and how the patient expects to be treated	'a crack in the bone' 'a broken bone' 'a fracture' 'a hairline fracture' 'a greenstick fracture'	Expected treatment (choice between heals on own, sling, cast, operation)	Perceived severity on a 10-point scale (1 being minimally problematic to 10 being a very serious injury)

*=included qualitative study

Table 3. Risk of bias summary for quantitative studies

Study	Study design	Study setting	Selection bias	Performance bias	Attrition bias	Reporting bias	Measurement Bias	Analysis
McCaffery, 2015	+	+	+	?	+	+	+	+
Omer, 2013	+	?	+	?	?	+	+	+
Scherer, 2013	+	+	+	?	?	+	+	+
Scherer, 2015	+	+	+	?	?	+	+	+
Azam, 2010	-	+	-	?	?	-	-	-

Table 4. Key findings from individual studies

Study	Primary Outcome	Primary Outcome by Terminology		Primary Outcome Statistical Significance*	Other Outcomes
		Medicalised term	Non-medicalised term		
McCaffery, 2015 (n=269)	Treatment preference	Pre-invasive breast cancer cells (n=128) 40% (51) prefer treatment 60% (77) prefer watchful waiting Change in terminology 41% (52) prefer treatment 59% (76) prefer watchful waiting	Abnormal cells (n=141) 33% (47) prefer treatment 67% (94) prefer watchful waiting Change in terminology 45% (63) prefer treatment 55% (78) prefer watchful waiting	Primary outcomes not significant however change in terminology shows a significant difference. There were no significant difference in treatment between arm A (women who were given the term abnormal cells first and then were given the term pre-invasive cancer cells) and arm B (women who were given the term pre-invasive cancer cells first and then were given the term abnormal cells) with 33% and 41% of women respectively favouring treatment, p=0.23. In arm A, 18% of women changed their preference to treatment while only 6% changed to watchful waiting (p=0.008). No significant treatment preferences were observed in arm B (9% vs 8%, p>0.99).	Level of concern: There was no significant difference between arms with 49% and 44% of women indicating they would be extremely concerned p=0.600. However, when the alternative term was used women in arm A (abnormal cells terminology first and then pre-invasive cancer cells terminology) were significantly more likely to report increased concern than women in arm B (pre-invasive cancer cells terminology first and then abnormal cells terminology) 67% vs 52%, p=0.001.
Cher, 2013 (n=394)	Treatment preference	Cancer 47% (186) surgery 20% (79) medication 33% (129) active surveillance	Lesion 34% (136) surgery 18% (70) medication 48% (188) active surveillance Abnormal cells 31% (124) surgery 21% (82) medication 48% (188) active surveillance	Primary outcomes significant (p=<.001) When DCIS was described as a non-invasive cancer 53% (208 of 394) of participants preferred non-surgical options, whereas 66% (258 of 394) chose non-surgical options when it was described as breast lesion and 69% (270 of 394) chose non-surgical options when it was described as abnormal cells (p=<.001).	
Scherer, 2013 (n=175)	Parents' interest in medicating their infant (3 items pooled and mean reported)	GERD term (n=87) mean=2.51 (on a 6-point Likert scale where 0=No, definitely not and 5=Yes, definitely)	No label (n=88) mean = 2.04 (on a 6-point Likert scale where 0=No, definitely not and 5=Yes, definitely)	Primary outcome significant (p<.01) Parents who received the GERD term in the scenario were more interested in medication than parents who did not receive that term, F(1,165) = 6.95, p<.01. Parents not given the GERD term were interested in medication only when they were not given information about medication effectiveness	Perception of illness severity: Findings were not influenced by the presence or absence of the GERD term (all p>.12). Appreciation of medication offer: Parents were least appreciative of medication when they were told that there medication was ineffective and were not given the GERD

5				therefore they were allowed to assume that the medications are effective $F(1,165) = 4.52, P<.05$.	label $F(1,165)=7.16, p<.01$
6					
7	Scherer,	Parents'	Pink-eye (n=82)	Eye infection (n=76)	Beliefs about contagiousness: Parents who
8	2015	decision to	mean=2.43	mean=2.32	received the 'pink-eye' term thought that
9	(n=158)**	give their	(on a 6-point Likert scale where	(on a 6-point Likert scale where	the symptoms were significantly more
10		child	0=No, definitely not and 5=Yes,	0=No, definitely not and 5=Yes,	contagious than parents who received the
11		antibiotics	definitely)	definitely)	'eye-infection' term, $F(1, 137)=11.21,$
12					$p=.001$.
13				Parents were less willing to give their child	Feelings towards sending child to day care:
14				antibiotics when symptoms were referred to as an	Parents who received the 'pink-eye' term
15				'eye infection' information about antibiotic	were less likely to believe that their child
16				ineffectiveness significantly reduced $F(1, 62) = 14.67,$	would be allowed to go to child care than
17				$p<.001$. By contrast when parents were told that	parents who received the 'eye-infection'
18				the symptoms were 'pink eye' interest in antibiotics	term, $F(1, 141)=9.70, p=.002$.
19				was not reduced by information about antibiotic	
20				ineffectiveness $F(1,74)=0.93 p=.33$.	
21	2010	Expected	Broken bone	Crack in the bone	Patients perspectives of severity:
22	(n=100)	treatment	3% heals on own	24% heals on own	(mean/median score, no SDs given):
23			39% sling	57% sling	3.28/3 for 'a crack in the bone'
24			56% cast	13% cast	3.58/3 for 'a hairline fracture'
25			2% operation	6% operation	4.95/5 for 'a fracture'
26					5.28/5 for 'a greenstick fracture'
27			Greenstick fracture	Hairline fracture	6.64/7 for 'a broken bone'
28			47% heals on own	45% heals on own	
29			25% sling	29% sling	
30			22% cast	26% cast	
31			6% operation	0% operation	
32					
33			Fracture		
34			14% heals on own		
35			44% sling		
36			36% cast		
37			6% operation		

*as reported by original study authors

**=1 case from total sample missing

Table 5. Summary of preferences for more invasive treatment option by type of terminology*

Study:	More medicalised term/s (%)	Less/non-medicalised term/s (%)	p-value
McCaffery, 2015	40	33	0.23
Omer, 2013	47	32.5	<0.001
Scherer, 2013**	74	67	.346§
Scherer, 2015**	60	58	.812§
Azam, 2010	42 (5 operation, 38 cast)†	22 (3 operation, 20 cast)‡	<0.025§

*=combined data where applicable and mean percentages reported
**=data from Likert scale with anchored end-points (0=No, definitely not; 5=Yes, definitely), assumed 1=No, 2=Maybe not, 3=Maybe, 4=Yes; therefore 0-1=non-invasive treatment preference, 2-5=invasive treatment preference
†=broken bone, green-stick fracture, fracture; ‡=hairline fracture, crack in the bone
§=calculated significance using raw (Scherer) and published (Azam) data based on our classification of which terms were more medicalised

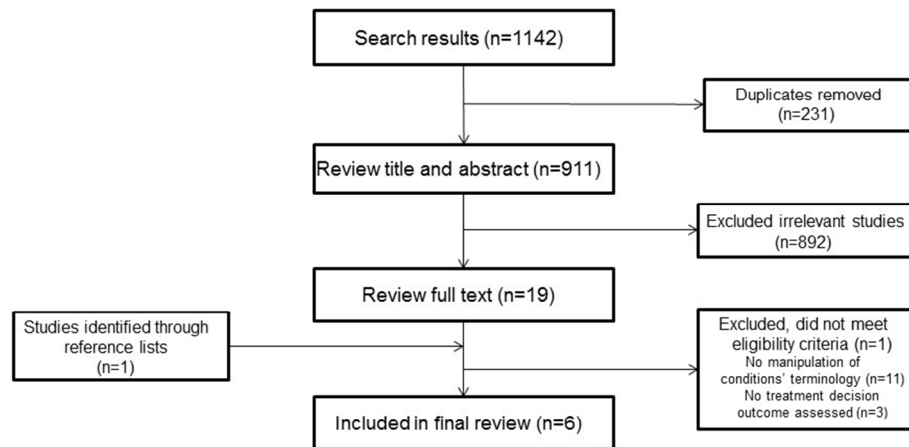


Figure 1. Flow diagram of study selection

254x190mm (96 x 96 DPI)

Appendix 1. Search strategy

OVID Medline	OVID Pre-Medline	Embase	PsycINFO	CINAHL	PubMed
1. exp Terminology as Topic/ 2. terminolog*.tw. 3. medical term*.tw. 4. disease label*.tw. 5. (disease adj3 label*).tw. 6. medical label*.tw. 7. 1 or 2 or 3 or 4 or 5 or 6 8. Patient Preference/ 9. Decision Making/ 10. patient decision making.tw. 11. (patient adj3 decision*).tw. 12. treatment decision making*.tw. 13. (treat* adj3 decision*).tw. 14. treatment pref*.tw. 15. (treat* adj3 pref*).tw. 16. management pref*.tw. 17. (manage* adj3 pref*).tw. 18. overdiagnosis.kw. 19. overtreatment.kw. 20. 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15 or 16 or 17 or 18 or 19 21. 7 and 20	1. terminolog*.tw. 2. medical term*.tw. 3. disease label*.tw. 4. (disease adj3 label*).tw. 5. medical label*.tw. 6. 1 or 2 or 3 or 4 or 5 7. patient preference/ 8. patient decision making.tw. 9. treatment decision making*.tw. 10. (treat* adj3 decision*).tw. 11. treatment pref*.tw. 12. (treat* adj3 pref*).tw. 13. management pref*.tw. 14. (manage* adj3 pref*).tw. 15. overdiagnosis.kw. 16. overtreatment.kw. 17. 7 or 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15 or 16 18. 6 and 17	1. nomenclature/ 2. terminolog*.tw. 3. medical term*.tw. 4. disease label*.tw. 5. (disease adj3 label*).tw. 6. medical label*.tw. 7. 1 or 2 or 3 or 4 or 5 or 6 8. patient preference/ 9. patient decision making/ 10. treatment decision making*.tw. 11. (treat* adj3 decision*).tw. 12. treatment pref*.tw. 13. (treat* adj3 pref*).tw. 14. management pref*.tw. 15. (manage* adj3 pref*).tw. 16. overdiagnosis.kw. 17. overtreatment.kw. 18. 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15 or 16 or 17 19. 7 and 18	1. terminology/ 2. terminolog*.tw. 3. medical term*.tw. 4. disease label*.tw. 5. (disease adj3 label*).tw. 6. medical label*.tw. 7. 1 or 2 or 3 or 4 or 5 or 6 8. decision making/ or management decision making/ 9. patient decision making.tw. 10. treatment decision making*.tw. 11. (treat* adj3 decision*).tw. 12. treatment pref*.tw. 13. (treat* adj3 pref*).tw. 14. management pref*.tw. 15. (manage* adj3 pref*).tw. 16. overdiagnosis.kw. 17. overtreatment.kw. 18. 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15 or 16 or 17 19. 7 and 18	S1 TI terminolog* or AB terminolog* S2 TI medical term* or AB medical term* S3 TI disease label* or AB disease label* S4 TI disease N3 label* or AB disease N3 label* S5 TI medical label* or AB medical label* S6 S1 OR S2 OR S3 OR S4 OR S5 S7 (MH "Decision Making") OR (MH "Decision Making, Patient") S8 TI treatment decision making or AB treatment decision making S9 TI treatment N3 decision* or AB treatment N3 decision* S10 TI treatment pref* or AB treatment pref* S11 TI treatment N3 pref* or AB treatment N3 pref* S12 TI management pref* or AB management pref* S13 TI manage* N3 pref* or AB manage* N3 pref* S14 "overdiagnosis" S15 "overtreatment" S16 S7 OR S8 OR S9 OR S10 OR S11 OR S12 OR S13 OR S14 OR S15 S17 S6 AND S16	(((((terminolog*[Title/Abstract]) OR medical term*[Title/Abstract]) OR medical label*[Title/Abstract]) OR disease label*)) AND ((((((treatment decision making[Title/Abstract]) OR treatment decision*[Title/Abstract]) OR treatment pref*[Title/Abstract]) OR management pref*[Title/Abstract]) OR overdiagnosis) OR overtreatment))

Appendix 2. Eligibility criteria (with exclusion criteria)

Types of studies:	Empirical studies that considered treatment decision making as an anticipated or experienced outcome of being given or told one terminology vs another for the same condition including surveys, questionnaires and interviews or focus groups <i>Exclusion: review papers, editorials, commentary/discussion papers</i>
Types of participants/population:	Adults (patients, community members) aged 18 years and above including adults making surrogate decisions or decisions on behalf of their children <i>Exclusion: patients or community members less than 18 years of age making a decision for themselves</i>
Types of settings:	Any type of medical or community setting including hypothetical scenarios given to community members
Study factor (intervention):	Different term given for the same condition
Outcome factor:	Treatment or management preferences (eg. treatment 1 vs treatment 2 vs treatment 3/no treatment)

Appendix 3. Nickel et al. qualitative study appraisal*

Were the researcher characteristics described?	Partly/moderate quality – interviews were conducted by two public health researchers with experience in qualitative research methods, however credentials, gender and characteristics of the interviewer were not reported
Was the methodological orientation to the study described?	Yes/ high quality – rationale informed by previous published DCIS terminology research (Esserman 2013, Omer 2013)
Was the target population clearly defined?	Yes/high quality – community sample of 26 Australian women aged 25 years and above who were recruited from a national community survey. Table 1 includes participants characteristics (age, education, experience with cancer screening, previous breast cancer diagnosis, immediate family history of cancer, employment status, worked as a health professional, survey management preference)
Was the sampling strategy clearly defined?	Yes/high quality – consenting women were recruited from a national community survey and purposively selected according to their education background (lower vs higher education), previous screening experience (previously screened vs not) and management preferences (immediate treatment vs watchful waiting as indicated by their survey responses)
Were the data collection methods clearly described?	Yes/high quality – semi-structured single telephone interview which lasted 13-40 minutes, and were audio-recorded and transcribed verbatim
Was data analysis clearly described?	Yes/high quality – Framework Analysis was used to organise the data and identify main themes that capture the diverse views expressed with two researches independently coding data using a rigorous analysis process

*criteria adapted from Tong et al.¹²

Appendix 4. Justification for treatment and terminology classification in each study

Study	Treatment	Classification	Justification	Terminology	Classification	Justification
McCaffery, 2015*	Treatment	Invasive treatment		Pre-invasive breast cancer cells	More medicalised term	Terminology which included the cancer term was deemed to be more medicalised than without.
	Watchful waiting	Non-invasive treatment		Abnormal cells	Less/non-medicalised term	
Omer, 2013*	Surgery	Invasive treatment		Cancer	More medicalised term	Terminology which included the term cancer was deemed to be more medicalised than those without.
	Medication Active surveillance	Non-invasive treatment		Lesion Abnormal cells	Less/non-medicalised term	
Scherer, 2013*	2-5 on Likert scale "Will you give your infant this medication?"	Invasive treatment	Likert scale with anchored end-points (0=No, definitely not; 5=Yes, definitely), assumed 1=No, 2=Maybe not, 3=Maybe, 4=Yes†	Gastroesophageal reflux disease (GERD)	More medicalised term	The condition's full medical terminology (GERD) was deemed to be more medicalised.
	0+1 on Likert scale "Will you give your infant this medication?"	Non-invasive treatment		"This condition" (no label)	Less/non-medicalised term	
Scherer, 2015*	2-5 on Likert scale "Will you give your child antibiotics?"	Invasive treatment	Likert scale with anchored end-points (0=No, definitely not; 5=Yes, definitely), assumed 1=No, 2=Maybe not, 3=Maybe, 4=Yes†	Pink-eye	More medicalised term	Giving the condition a specific terminology (pink-eye) rather than a generic term (eye infection) was deemed to be more medicalised.
	0+1 on Likert scale "Will you give your child antibiotics?"	Non-invasive treatment		Eye infection	Less/non-medicalised term	
Azam, 2010	Operation Cast	Invasive treatment		Broken bone Fracture Greenstick fracture	More medicalised term	Authors determined classification based on the severity rating participants gave each term in the study. Other literature on perceptions of injury terminology was searched by authors with no results.
	Heals on own Sling	Non-invasive treatment		Crack in the bone Hairline fracture	Less/non-medicalised term	

*=studies where authors were specifically guided by study design to classify terminology

†=sensitivity analysis conducted



PRISMA 2009 Checklist

Section/topic	#	Checklist item	Reported on page #
TITLE			
Title	1	Identify the report as a systematic review, meta-analysis, or both.	1
ABSTRACT			
Structured summary	2	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.	2-3
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known.	5
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).	6
METHODS			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.	6
Eligibility criteria	6	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	7 (Appx.2)
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	6-7
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	6-7 (Appx.1)
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	7
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	6-7
Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.	6-7
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.	7 (Table 1 & Appx. 3)
Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means).	8
Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., I^2 for each meta-analysis)	8 (Appx. 4)



PRISMA 2009 Checklist

Page 1 of 2

Section/topic	#	Checklist item	Reported on page #
Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).	N/A
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.	
RESULTS			
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.	8
Study characteristics	18	For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.	8 (Table 2)
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12).	8-9 (Table 3)
Results of individual studies	20	For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot.	9-13 (Table 4)
Synthesis of results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.	13 (Table 5)
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see Item 15).	N/A
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]).	
DISCUSSION			
Summary of evidence	24	Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy makers).	13-14
Limitations	25	Discuss limitations at study and outcome level (e.g., risk of bias), and at review-level (e.g., incomplete retrieval of identified research, reporting bias).	14-16
Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.	16
FUNDING			
Funding	27	Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review.	17

BMJ Open

Words do matter: a systematic review on how different terminology for the same condition influences management preferences

Journal:	<i>BMJ Open</i>
Manuscript ID	bmjopen-2016-014129.R1
Article Type:	Research
Date Submitted by the Author:	23-Apr-2017
Complete List of Authors:	Nickel, Brooke; University of Sydney, School of Public Health Barratt, Alexandra; University of Sydney, School of Public Health Copp, Tessa; University of Sydney School of Public Health, Moynihan, Ray; Bond University , Centre for Research in Evidence-Based Practice McCaffery, Kirsten; The University of Sydney, Screening and Test Evaluation Program (STEP), School of Public Health; The University of Sydney, Centre for Medical Psychology & Evidence-based Decision-making (CeMPED)
Primary Subject Heading:	Communication
Secondary Subject Heading:	Medical management
Keywords:	terminology, disease labels, management preferences, overtreatment, medical decision making

SCHOLARONE™
Manuscripts

only

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Words do matter: a systematic review on how different terminology for the same condition influences management preferences

Brooke Nickel *PhD candidate*¹, Alexandra Barratt *epidemiologist*¹, Tessa Copp *PhD candidate*¹, Ray Moynihan *research fellow*^{1,2}, Kirsten McCaffery *health psychologist and professorial research fellow*¹

¹Wiser Healthcare, Sydney School of Public Health, The University of Sydney, Sydney, New South Wales, Australia

²Centre for Research in Evidence-Based Practice, Bond University, QLD 4229, Australia

Correspondence to: Ms Brooke Nickel; Wiser Healthcare, Sydney School of Public Health, The University of Sydney, NSW 2006, Australia. (email: brooke.nickel@sydney.edu.au)

ABSTRACT

Objectives: Changing terminology for low-risk, screen detected conditions has now been recommended by several expert groups in order to prevent overdiagnosis and reduce the associated harms of overtreatment. However, the effect of terminology on patients' preferences for management is not well understood. This review aims to synthesise existing studies on terminology and its impact on management decision making.

Design: Systematic review.

Methods: Studies were included that compared two or more terminologies to describe the same condition and measured the effect on treatment or management preferences and/or choices. Studies were identified via database searches from inception to April 2017, and from reference lists. Two authors evaluated the eligibility of studies with verification from the study team, extracted and cross-checked data, and assessed the risk of bias of included studies.

Results: Of the 1399 titles identified, 7 studies, all of which included hypothetical scenarios, met the inclusion criteria. Six studies were quantitative and one was qualitative. Six of the studies were of high quality. Studies covered a diverse range of conditions: ductal carcinoma in situ (3), gastroesophageal reflux disease (1), conjunctivitis (1), polycystic ovary syndrome (1) and a bony fracture (1). The terminologies compared in each study varied based on the condition assessed. Generally, when a more medicalised or precise term was used to describe the condition it resulted in a shift in preference towards more invasive managements, and/or higher ratings of anxiety and perceived severity of the condition.

Conclusions: Different terminology given for the same condition influenced management preferences and psychological outcomes in a consistent pattern in these studies. Changing the terminology may be one strategy to reduce patient preferences for aggressive management responses to low-risk conditions.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Registration: CRD42016035643.

For peer review only

STRENGTHS AND LIMITATIONS OF THIS STUDY:

- This is the first systematic review to synthesise the evidence on how different terminology given for the same condition impacts patients' management preferences
- Only a small number of studies have examined this research question and were included in the review
- Due to the variability of terms and outcomes assessed, authors were unable to conduct a meta-analysis and pool the effects of the data
- All studies included were hypothetical, therefore patients facing a real diagnosis may respond differently

INTRODUCTION

Medical encounters can be challenging and confronting for patients, especially when they are faced with a management decision. Clinical communication and language is an important aspect of a medical encounter as it influences patients’ understanding of their diagnosis and management options^{1,2}. Decisions about treatments or tests may be influenced by various factors including the medical terminology clinicians use to diagnose and describe conditions to patients. Describing a patients’ circumstances using more medicalised terminology may lead to greater preference for more invasive managements, whether or not the extent of those treatments or tests are warranted.

Overdiagnosis of several medical conditions and associated overtreatment is now widely accepted^{3,4}, and can have serious implications for patients, healthcare systems and society^{5,6}. Numerous approaches are beginning to be proposed to help combat overdiagnosis and overtreatment, including various communication strategies⁷. Changing the terminology for medical conditions may be one communication strategy to mitigate the effect of overdiagnosis and overtreatment as it has the potential to influence the effect of diagnosis and labelling of a condition, and influence patients’ decision making about management. It may encourage both patients and clinicians to more carefully consider conservative management options.

In particular, cancer terminology is one area where use of different terminology may greatly influence management decision making. The term ‘cancer’ is understandably frightening for people to hear and can influence their thought and action⁸, but it is now well understood that a range of conditions which include indolent to fast-growing lesions are labelled as cancer⁹. Lesions with low malignant potential are common (such as low-risk DCIS, low-risk papillary thyroid cancer and low-grade prostate cancer), and with the advent and increasing use of various screening technologies these indolent lesions and their precursors are now frequently clinically identified. This identification can lead to the condition being overdiagnosed and in turn overtreated.

Changing the terminology for these low-risk, screen detected conditions has now been recommended by several expert groups – including a National Institutes of Health state of the science conference panel and a National Cancer Institute working group – in order to prevent overdiagnosis and encourage more consideration of less invasive management options such as active surveillance⁹⁻¹¹. However, the effect of terminology on patients' willingness to accept more conservative management options is not well understood. This systematic review aims to synthesise existing studies on terminology for medical conditions and its impact on management decision making, and associated psychosocial outcomes.

METHODS

Protocol and registration

The review's protocol is registered with PROSPERO (an international prospective register of systematic reviews), registration number: CRD42016035643.

Review Question

How do different terminologies given to the same condition influence treatment or management decision making?

Search Strategy

A comprehensive list of search terms was developed (Appendix 1) with consultation from an information specialist and a search of relevant databases (Medline, Pre-Medline, Embase, PsycINFO, Cinhal, and PubMed) was conducted from inception to April 2017. The returned search results were screened by title and abstract independently by two researchers (BN and TC) for irrelevant articles, reviews papers, editorials or commentaries, and duplicates. An eligibility checklist was developed (Appendix 2) to guide the selection of appropriate studies. Decisions regarding inclusion and exclusion of studies was then made independently by two researchers (BN and TC) and

disagreements discussed. Any further disagreement or uncertainty was discussed and verified by two additional researchers (AB and KM). A hand search of reference lists of included studies as well as papers recommended through personal communication were also examined for relevant studies.

Inclusion and Exclusion Criteria

Studies were included in the review if they empirically measured treatment or management decision making as an anticipated or experienced outcome of being given or told one terminology versus another or others for a specific condition. Studies were excluded if they were reviews, editorials or commentaries, or if they assessed participants less than 18 years of age making a decision for themselves (Appendix 2).

Quality Assessment and Data Extraction

All studies that met the inclusion criteria were appraised for study quality by two authors (BN and TC) independently using two separate tools; one for quantitative studies and one for the qualitative study. The quantitative studies were appraised using a modified version of the Cochrane Collaboration’s tool for assessing risk of bias which was developed by study authors (Table 1). As the quantitative studies included in this review were not clinical trials, study authors adapted the Cochrane Collaboration’s tool for assessing risk of bias to appraise the studies for items relating to study design, study setting, study validity, and analysis. Study validity included allocation of participants to different terminology, blinding of participants to the study hypotheses, and reporting and measurement bias. The qualitative study was appraised using criteria adapted from the Consolidated criteria for reporting qualitative studies (COREQ) framework¹² and results appear in Appendix 3.

Two categories of study quality were identified by study authors according to each study’s methodological characteristics. In high-quality studies (lower risk of bias) the majority of criteria were fulfilled and done well, while in low-quality studies (higher risk of bias) the majority of criteria

were not done or done poorly. Results of the risk of bias assessment for quantitative studies are shown in Table 3.

Data from the final studies included in the review were extracted independently into a standardised template by two authors (BN and TC). Any discrepancies were discussed and resolved by the entire research team. For studies where the data was not reported or was unclear, authors were contacted and responded for confirmation of details of the data. Authors were also asked to confirm the extracted data to minimise any reporting bias. All authors responded. Results from the studies were synthesised in a narrative form, as the heterogeneity of the studies and their respective outcome measures did not support pooling of results¹³.

RESULTS

Initial search results identified 1399 papers. After removal of duplicates and screening by title and abstract, 20 studies from the search and 1 study identified from reference lists underwent full-text reviews. Seven studies that reported the impact of different terminology for the same condition on treatment or management decision making were identified (Figure 1). Three studies reported on ductal carcinoma in situ (DCIS)¹⁴⁻¹⁶, two on common childhood conditions (gastroesophageal reflux disease (GERD), conjunctivitis)^{17,18}, one on polycystic ovary syndrome (PCOS)¹⁹ and one on a bony fracture²⁰. Six studies reported quantitative findings^{14,15,17-20} and one study reported qualitative findings¹⁶. The key characteristics of these studies and their methods are summarised in Table 2. All studies were hypothetical and involved various samples of community members who were not currently and/or previously diagnosed with the condition assessed. Of the six quantitative studies, five involved a randomised experimental design; with two using a paired sample design (cross-over) and three using an independent sample design (2x2 factorial). The qualitative study was linked to one of the quantitative studies¹⁵, as women from the study were asked if they would be willing to participate in an additional qualitative interview for further investigation of the topic.

Overall the majority of studies were of higher quality with a lower risk of bias, however one study did have a high risk of bias as the quality of reporting was low²⁰ (Table 3). Higher quality studies had defined study populations and settings, had low selection bias and described the data collection, methods and analysis appropriately. The qualitative study included in the review was rigorous and had appropriate methods to reach its research objectives. Researchers and the target population were defined, the sampling strategy was explained, data collection methods and data analysis were appropriate and well documented.

Since the strength of the review is the diversity of included studies, results of the studies are summarised individually by study quality category, with data on management outcomes summarised in Table 4. Following this is a narrative synthesis of the effect of terminology (more medicalised or precise terms vs less or non-medicalised terms) on management preferences (invasive management preference vs non-invasive management preference). Importantly for each study we identify the classifications of management preferences and terminology and provide justification where applicable (Appendix 4).

Results from individual higher quality studies

Copp et al. 2017 (Polycystic ovary syndrome)

This study on the influence of the term PCOS found that when young women were given the term ‘PCOS’ in a hypothetical scenario of a doctor’s visit, they had significantly higher intention to have an ultrasound compared to women who were given the term ‘hormonal imbalance’ (mean=6.62 vs. mean=5.76, $F(1, 176)=4.63$, $p=0.033$). The study also found that those who received the ‘PCOS’ term perceived their hypothetical condition to be more severe (17.7 vs. 15.82, $F(1, 176)=5.65$, $p=0.019$) and had lower self-esteem compared to women who were not given the PCOS term (25.86 vs. 27.56, $F(1, 176)=4.74$, $p=0.031$). After women received information about the potential of PCOS

overdiagnosis in a second scenario, both intention and perceived severity decreased, regardless of the term given (both $p < 0.001$).

The study also found a significant 3-way interaction between the term 'PCOS', information about ultrasound reliability and overdiagnosis information ($F(1, 176) = 4.23$, $p = 0.041$), where that for those who did not receive the PCOS term, intention was significantly lower for those who received information about the unreliability of ultrasounds compared with those who received no information. For women who received the PCOS term however, intention was high, even when told the ultrasound was unreliable. This difference disappears after information about overdiagnosis is given in the second scenario, suggesting the provision of overdiagnosis information reduces the effect of the term.

McCaffery et al. 2015 (Ductal carcinoma in situ)

McCaffery and colleagues' study on the impact of DCIS terminology on treatment preference (immediate treatment vs watchful waiting) and women's level of concern found no significant differences in treatment preference between arm A (women who were given the term 'abnormal cells' first and then were given the term 'pre-invasive cancer cells') and arm B (women who were given the term 'pre-invasive cancer cells' first and then were given the term 'abnormal cells') with 33% and 40% of women respectively favouring treatment ($p = 0.23$). However, 18% of women in arm A who were initially given the 'abnormal cells' terminology changed their preference to treatment when the terminology was switched to 'pre-invasive breast cancer cells' while only 6% changed to watchful waiting ($p = 0.008$). In contrast, there were no significant changes in treatment preference in arm B when the terminology was switched the other way (9% vs 8% changed their stated preference).

Similarly, this study found that there was no significant difference between arms with regards to level of concern with 49% and 44% of women indicating they would be extremely concerned

p=0.600. However, when the alternative term was used, women in arm A (who were initially given the ‘abnormal cells’ term and then were given the ‘pre-invasive breast cancer cells’ term) were significantly more likely to report increased concern than women in arm B (67% vs 52%, p=0.001). Findings from this study were also supported by the included qualitative study conducted with a subset of women¹⁶.

Nickel et al. 2015 (Ductal carcinoma in situ)

This qualitative study investigated in-depth how different proposed terminologies to describe DCIS affected women’s treatment preferences and psychological outcomes by conducting semi-structured interviews with women of varying education, cancer screening experience and with no history of a DCIS diagnosis. Findings demonstrated that overall women preferred a diagnosis of DCIS to be communicated using terminology that did not include the term cancer, as women generally exhibited stronger negative reactions when the cancer term was used to describe DCIS compared to when a non-cancer term was used. Although concern seemed to be high overall, women displayed a high level of interest in watchful waiting when it was described to them in a hypothetical scenario as a safe and effective option, and told that they could proceed to treatment in the future if necessary.

Omer et al. 2013 (Ductal carcinoma in situ)

This study also examined the impact of DCIS terminology on women’s treatment preferences by comparing 3 identical scenarios, with the only difference being the term used to described DCIS (‘non-invasive cancer’, ‘breast lesion’ and ‘abnormal cells’). Treatment options presented were surgery, medication, active surveillance. All participants saw all scenarios, with the order of scenarios varied across participants. When DCIS was described as a non-invasive cancer 53% of participants preferred non-surgical options, whereas 66% chose non-surgical options when it was described as breast lesion and 69% chose non-surgical options when it was described as abnormal cells (p<0.001). Although women with a previous history of cancer (other than breast cancer) and

women with high socioeconomic status more frequently chose surgery in univariate analyses, high numeracy was the only independent predictor of preference for surgical treatment in the multiple variable logistic regression model for all three terms: cancer (OR 2.11, 1.34-3.34 CI, $p=0.001$), lesion (OR 1.96, 1.20-3.19, $p=0.001$), abnormal cells (OR 1.63, 1.01-2.67, $p=0.048$).

Scherer et al. 2013 (Gastroesophageal reflux disease)

This study on the influence of the term 'GERD' (versus no term or label given) on parents preferences for medication for their infant found that parents who received the term GERD in the scenario were more interested in medication than parents who did not receive that term, ($F(1, 165)=6.95$, $p<0.01$). To assess parental interest in antibiotics, study authors combined three highly inter-correlated questions involving parent's interest in medication: Will you give your infant this medicine? Do you think your infant needs the medicine your doctor offered? Do you think that the medicine will help your infant get better?

The study also found a significant interaction between the term GERD and report of medicine ineffectiveness, ($F(1, 165)=4.52$, $p<0.05$) as parents who received a GERD diagnosis were interested in medicating the infant, even if they were told that the medications are likely ineffective. By contrast parents not given a diagnosis were interested in medication only when they were not given information about medication effectiveness, thus able to assume that the medications are effective.

All parents in the study were also asked (using a 0-5 Likert scale with labelled end-points for each question) whether they were worried about their infant's health (mean=2.28, SD=1.30), thought the condition was somewhat serious (mean=2.12, SD=1.19), and were relatively unlikely to describe their infant as being sick (mean=1.87, SD=1.45). These answers were not influenced however by the presence or absence of the GERD term (all $p>0.12$).

Scherer et al. 2015 (Conjunctivitis)

A similar study design conducted by Scherer and colleagues did not find an initial difference between the term 'pink-eye' and 'eye-infection' on parents' preference to medicate their infant. However, when symptoms were referred to as an 'eye infection', information about antibiotic ineffectiveness significantly reduced interest in using medication ($F(1, 62) = 14.67, p < 0.001$). By contrast, when parents were told that the symptoms were 'pink eye', interest in antibiotics was not reduced by information about antibiotic ineffectiveness ($F(1, 74) = 0.93, p = 0.33$).

Study authors also measured (using a 0-5 Likert scale with labelled end-points) parents' perceptions about contagiousness and belief that their child could attend child care and found that parents who received the 'pink-eye' term thought that the symptoms were significantly more contagious (mean=4.04, SD=1.30) than parents who received the 'eye-infection' term (mean=3.29, SD=1.63), ($F(1, 137) = 11.21, p = 0.001$). Parents who received the 'pink-eye' term were also less likely to believe that their child would be allowed to go to child care (mean=0.35, SD=0.90) than parents who received the 'eye-infection' term (mean=0.93, SD=1.13), ($F(1, 141) = 9.70, p = .002$).

Individual results from the lower quality study

Azam et al. 2010 (Bony Fracture)

This study found that patients' management expectations and perceptions of bony injuries differ based on the terminology used to describe the injury. Patients expected more invasive managements (operation or cast) when a more medicalised term was used to describe the injury compared to a less invasive treatment (sling or heals on own), with 58% of patients expecting invasive treatments for the term 'broken bone', 42% for 'fracture', 28% for 'greenstick fracture' and 26% for hairline fracture. In contrast, patients given the less medicalised term had a lower preference for invasive treatments, with 19% of patients choosing an invasive treatment for 'crack in the bone'. This study also found that patients perceived the injury to be more severe when a more

medicalised term was used to describe the injury (average mean severity score out of 10; broken bone=6.64, greenstick fracture=5.28, fracture=4.95, hairline fracture=3.58, crack in the bone=3.28).

Synthesis of results

Table 5 summarises preferences for more invasive management option by type of terminology across the studies. Generally, there was a pattern in the same direction showing that when more medical or precise term was used it resulted in a greater preference or interest in more invasive managements, whether this be intention to have an ultrasound for a PCOS diagnosis, surgical treatment for DCIS, increased interest in potentially ineffective medication, or an operation or cast for a bony injury, although these differences did not always reach statistical significance. In contrast, when participants were given a less or non-medical/precise term a higher proportion of participants chose a non-invasive management option. In this analysis medicalised or precise terminology refers to language that describes the condition either using medical terms that healthcare professionals commonly used or that described the condition in a more specific way (when compared to the comparator term). Studies varied as some found a significant effect of terminology on management decision making while others found a significant interactions effect or within person effect and psychological outcome effects (e.g. perceived severity, level of concern).

DISCUSSION

This is the first systematic review which synthesises the evidence on how different terminology given for the same condition impacts management preferences. Overall, the review demonstrates that different terminology used to describe the same condition can influence patient's management preferences. Results indicate that when a more medical or more precise term was used to describe a condition, people tended to have stronger preferences for more invasive management options. Although not all of the studies included in our review had results which were statistically significant in relation to preferences for more invasive managements, at a population level these trends may

represent a clinically important difference. For example, a relatively small increase in the number of people preferring surgery in these studies could translate into significantly more surgeries across a larger population.

The terminology given during a diagnosis to patients is important, as when a healthy individual becomes a patient they can immediately become more vulnerable to the words their clinicians use²¹. Using words that generate fear or anxiety cause patients to have more difficulty making informed decisions and becoming an active participant in their care²². Different terms used to describe the same condition can have a direct influence on how patients understand their diagnosis and how threatening they perceive it to be²³. It has also been shown that the use of interpretive terminology (e.g. including the words positive or negative, or using a metaphor)^{24 25}, the terminology used to describe a treatment choice²⁶, describing a condition with plain language terminology as compared to jargon²⁷ and the severity of the characteristics of the diagnosis²⁸ can have an influence on medical decision making. Furthermore, a recent survey of medical students found that students were more likely to classify synonyms as a 'disease' if the term used to describe it was medical²⁹. Together with findings from this review, these studies show that language is a powerful tool that has the potential to influence patients' thoughts and actions.

Our findings are also in line with the results of other research on the effect of labelling a condition in relation to the social implications it may have for the individual, rather than its effect on medical decision making. Research has found that disease labels can result in various emotional, cognitive and physical consequences^{30 31}. Most notably, studies which have examined the effect of hypertension labelling have found that giving the label of hypertension to individuals (compared to giving no label) resulted in increased self-reported illness and absenteeism from work, and a significant increase in blood pressure^{30 32 33}. In the conjunctivitis study by Scherer and colleagues included in this review, the term 'pink-eye' was perceived as being more contagious, and parents were less likely to believe their child could go to childcare, compared with parents who received the

condition described as an 'eye-infection'. Furthermore, in the PCOS study women had significantly lower levels of self-esteem when the term 'PCOS' was used to describe their condition compared to when it was described as a hormonal imbalance¹⁹.

Using more medicalised or precise medical terminology and disease labels may also reduce a patient's sense of self-control, and therefore increase perceived severity and uptake of medications. The more medical the term or label, the less control a person may feel over the situation, increasing their perceived severity about the condition and creating a perception that more invasive interventions are warranted. This was also shown in the GERD, conjunctivitis and PCOS study where there was a high level of interest in medical intervention when a precise medical term or label was given, even when participants were told that the medication or test was ineffective^{17 18}.

This study was limited by the small number of included studies in the review, which also included three studies that were conducted by members of the review team. However, in light of recent and ongoing evidence of overtreatment across a number of conditions and suggestions from leading global medical bodies to change the terminology of low-risk conditions^{10 11}, this review addressed an important and timely under-researched question.

Changing the terminology of low-risk conditions may be difficult in practice as a systems level approach would need to be taken to ensure that all healthcare professionals implemented the new terminology. Although, it would be feasible as demonstrated by the recent change to the terminology of the 'non-invasive encapsulated follicular variant of papillary thyroid carcinoma' (EFVPTC) to be 'non-invasive follicular thyroid neoplasm with papillary-like nuclear features' (NIFTP) in order to highlight the true nature of the tumour, lessen the emotional and psychological burden associated with the term 'cancer' and potentially reduce overtreatment^{34 35}.

Due to the variability of the terms and outcomes assessed in the included studies, authors were unable to conduct a meta-analysis and pool the effects of the data, and therefore a definite

synthesis of results of all studies was not possible. The delineation between more medical/precise and less or non-medical/precise was challenging in some studies. For example, it seemed clear that ‘pre-invasive breast cancer cells’ was a more medicalised term than ‘abnormal cells’. In contrast, it was not as clear what it was exactly about the term ‘pink eye’ (a more precise term to describe the condition) that elicited stronger parental preferences for antibiotics when compared with the term ‘eye infection’. Other aspects important to parents in this study (e.g. aesthetical aspects), may have influenced management preferences. The author team therefore made explicit judgements about which terms were deemed more medicalised or precise and which were not, as well as what managements were considered invasive and what were considered non-invasive. These decisions were guided from the aims and outcomes of the studies, and followed categories used in the original studies, with the exception of the Azam paper where authors were guided by information on the precise medical terminology healthcare professionals use to describe a bony injury (including ‘broken bone’, ‘fracture’, ‘greenstick fracture’, ‘hairline fracture’) (see Appendix 4). We note the Azam paper was of lower quality which made interpretations about the terms used more difficult. Nonetheless, it was very clear that in all of the studies the use of different terminology for the same condition tends to elicit different responses to management preferences as well as psychosocial outcomes. Further research could usefully explore in more detail what characterises terms that elicit stronger preferences for more invasive managements.

Importantly, all studies included in this review were hypothetical and were not presented to patients randomised to receive a diagnosis and management preferences were assessed instead of actual management decision making. Patients facing real decisions may respond differently to those in the studies, however it is likely that in real life situations these effects may be even more pronounced. It is likely that patients would be more anxious than participants in current studies, therefore more susceptible to judgement biases that result from more medicalised terminology or labels.

Against a backdrop of recent evidence of overdiagnosis and overtreatment across a number of conditions^{5 36-38} and the potential physical and psychological impacts this may have for the patient, understanding how different terminology given for the same medical condition may influence patients' management preferences is important. This review suggests that the terminology used to describe a condition can influence patient preferences for management and related outcomes. Although further evidence is needed to help better understand precisely why some terms elicited stronger preferences for more invasive management, this review helps support the calls for changing the terminology of conditions where the risk of progression is low. Changing the terminology in low-risk conditions or conditions with indolent clinical course may be a potential communication strategy to help shift assumptions that immediate invasive treatments or tests are always needed, allow for better shared decision making between clinicians and patients, and the consideration of more conservative management options.

Acknowledgements: The authors would like to thank Philippa Bourke for her help with the literature search and Dr Kevin McGeechan for his comments on the statistical aspects of the paper.

Contributions: All authors included in the paper fulfilled the criteria for authorship. BN contributed to study design, data collection, analysis, interpretation, drafting and revising the manuscript. AB contributed to study design, analysis, interpretation, and revising the manuscript. TC contributed to data collection, analysis, interpretation, and revising the manuscript. RM contributed to interpretation and revising the manuscript. KM contributed to study design, analysis, interpretation and revising the manuscript. All authors approved the final version of the manuscript and all authors are guarantors.

Funding: This paper was written with support from the Wiser Healthcare CRE grant from the National Health and Medical Research Council (NHMRC) #1104136. BN was supported by the Sydney

Catalyst Research Scholar Award. KM was supported by an NHMRC fellowship (1029241). The funders had no role in the design or conduct of the study; in the collection, analysis, and interpretation of the data; or in the preparation or approval of the manuscript.

Competing interests: All authors have completed the ICMJE uniform disclosure form at www.icmje.org/coi_disclosure.pdf and declare: no support from any organisation for the submitted work; no financial relationships with any organisations that might have an interest in the submitted work in the previous three years; AB and RM are a co-chairs of the steering group for the *Preventing Overdiagnosis* scientific conferences; no other relationships or activities that could appear to have influenced the submitted work.

Data sharing: No additional data available.

REFERENCES

1. Schofield PE, Butow PN, Thompson JF, et al. Psychological responses of patients receiving a diagnosis of cancer. *Ann Oncol* 2003;**14**(1):48-56.
2. Arora NK. Interacting with cancer patients: the significance of physicians' communication behavior. *Soc Sci Med* 2003;**57**(5):791-806.
3. Hoffman JR, Cooper RJ. Overdiagnosis of disease: a modern epidemic. *Arch Intern Med* 2012;**172**(15):1123-4. doi: 10.1001/archinternmed.2012.3319
4. Welch HG, Schwartz LM, Woloshin S. Over-diagnosed: making people sick in the pursuit of health. Boston: Beacon Press 2011.
5. Moynihan R, Doust J, Henry D. Preventing overdiagnosis: how to stop harming the healthy. *BMJ* 2012;**344**:e3502. doi: 10.1136/bmj.e3502
6. Heath I. Overdiagnosis: when good intentions meet vested interests--an essay by Iona Heath. *BMJ* 2013;**347**:f6361. doi: 10.1136/bmj.f6361
7. McCaffery KJ, Jansen J, Scherer LD, et al. Walking the tightrope: communicating overdiagnosis in modern healthcare. *BMJ* 2016;**352**:i348. doi: 10.1136/bmj.i348
8. Dunn BK, Srivastava S, Kramer BS. The word "cancer": how language can corrupt thought. *BMJ* 2013;**347**:f5328. doi: 10.1136/bmj.f5328
9. Esserman LJ, Thompson IM, Reid B, et al. Addressing overdiagnosis and overtreatment in cancer: a prescription for change. *Lancet Oncol* 2014;**15**(6):e234-42. doi: 10.1016/S1470-2045(13)70598-9
10. Allegra CJ, Aberle DR, Ganschow P, et al. National Institutes of Health State-of-the-Science Conference statement: Diagnosis and Management of Ductal Carcinoma In Situ September 22-24, 2009. *J Natl Cancer Inst* 2010;**102**(3):161-9. doi: 10.1093/jnci/djp485
11. Esserman LJ, Thompson IM, Jr., Reid B. Overdiagnosis and overtreatment in cancer: an opportunity for improvement. *JAMA* 2013;**310**(8):797-8. doi: 10.1001/jama.2013.108415

12. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. *Int J Qual Health Care* 2007;**19**(6):349-57. doi: 10.1093/intqhc/mzm042
13. Schünemann HJ, Oxman AD, Higgins JPT, et al. Presenting results and 'Summary of findings' tables — results without meta-analyses 2011.
14. Omer ZB, Hwang ES, Esserman LJ, et al. Impact of ductal carcinoma in situ terminology on patient treatment preferences. *JAMA Intern Med* 2013;**173**(19):1830-1. doi: 10.1001/jamainternmed.2013.8405
15. McCaffery K, Nickel B, Moynihan R, et al. How different terminology for ductal carcinoma in situ impacts women's concern and treatment preferences: a randomised comparison within a national community survey. *BMJ Open* 2015;**5**(11):e008094. doi: 10.1136/bmjopen-2015-008094
16. Nickel B, Barratt A, Hersch J, et al. How different terminology for ductal carcinoma in situ (DCIS) impacts women's concern and management preferences: A qualitative study. *Breast* 2015;**24**(5):673-9. doi: 10.1016/j.breast.2015.08.004
17. Scherer LD, Finan C, Simancek D, et al. Effect of "Pink Eye" Label on Parents' Intent to Use Antibiotics and Perceived Contagiousness. *Clin Pediatr (Phila)* 2015 doi: 10.1177/0009922815601983
18. Scherer LD, Zikmund-Fisher BJ, Fagerlin A, et al. Influence of "GERD" label on parents' decision to medicate infants. *Pediatrics* 2013;**131**(5):839-45. doi: 10.1542/peds.2012-3070
19. Copp T, McCaffery K, Azizi L, et al. Influence of the disease label 'polycystic ovary syndrome' on intention to have an ultrasound and psychosocial outcomes: a randomised online study in young women. *Hum Reprod* 2017;1-9. doi: 10.1093/humrep/dex029
20. Azam N, Harrison M. Patients' perspectives on injuries. *Emerg Med J* 2011;**28**(7):601-3. doi: 10.1136/emj.2009.082032
21. Bedell SE, Graboyes TB, Bedell E, et al. Words that harm, words that heal. *Arch Intern Med* 2004;**164**(13):1365-8. doi: 10.1001/archinte.164.13.1365
22. Cassell EJ, Leon AC, Kaufman SG. Preliminary evidence of impaired thinking in sick patients. *Ann Intern Med* 2001;**134**(12):1120-3.
23. Chadha NK, Repanos C. Patients' understanding of words used to describe lumps: a cross-sectional study. *J Laryngol Otol* 2006;**120**(2):125-8. doi: 10.1017/S0022215105004688
24. Zikmund-Fisher BJ, Fagerlin A, Keeton K, et al. Does labeling prenatal screening test results as negative or positive affect a woman's responses? *Am J Obstet Gynecol* 2007;**197**(5):528 e1-6. doi: 10.1016/j.ajog.2007.03.076
25. Scherer AM, Scherer LD, Fagerlin A. Getting ahead of illness: using metaphors to influence medical decision making. *Med Decis Making* 2015;**35**(1):37-45. doi: 10.1177/0272989X14522547
26. Rot I, Ogah I, Wassersug RJ. The language of prostate cancer treatments and implications for informed decision making by patients. *Eur J Cancer Care (Engl)* 2012;**21**(6):766-75. doi: 10.1111/j.1365-2354.2012.01359.x
27. Pieterse AH, Jager NA, Smets EM, et al. Lay understanding of common medical terminology in oncology. *Psychooncology* 2013;**22**(5):1186-91. doi: 10.1002/pon.3096
28. Gavaruzzi T, Lotto L, Rumiati R, et al. What makes a tumor diagnosis a call to action? On the preference for action versus inaction. *Med Decis Making* 2011;**31**(2):237-44. doi: 10.1177/0272989X10377116
29. Erueti C, Glasziou P, Mar CD, et al. Do you think it's a disease? a survey of medical students. *BMC Med Educ* 2012;**12**:19. doi: 10.1186/1472-6920-12-19
30. Pickering TG. Now we are sick: labeling and hypertension. *J Clin Hypertens (Greenwich)* 2006;**8**(1):57-60.

31. Hoyt CL, Burnette JL, Auster-Gussman L. "Obesity is a disease": examining the self-regulatory impact of this public-health message. *Psychol Sci* 2014;**25**(4):997-1002. doi: 10.1177/0956797613516981

32. Haynes RB, Sackett DL, Taylor DW, et al. Increased absenteeism from work after detection and labeling of hypertensive patients. *N Engl J Med* 1978;**299**(14):741-4. doi: 10.1056/NEJM197810052991403

33. Rostrup M, Kjeldsen SE, Eide IK. Awareness of hypertension increases blood pressure and sympathetic responses to cold pressor test. *Am J Hypertens* 1990;**3**(12 Pt 1):912-7.

34. Nikiforov YE, Seethala RR, Tallini G, et al. Nomenclature Revision for Encapsulated Follicular Variant of Papillary Thyroid Carcinoma: A Paradigm Shift to Reduce Overtreatment of Indolent Tumors. *JAMA Oncol* 2016 doi: 10.1001/jamaoncol.2016.0386

35. Patel KN. Noninvasive Encapsulated Follicular Variant of Papillary Thyroid "Cancer" (or Not): Time for a Name Change. *JAMA Oncol* 2016 doi: 10.1001/jamaoncol.2016.0714

36. Independent UK Panel on Breast Cancer Screening. The benefits and harms of breast cancer screening: an independent review. *Lancet* 2012;**380**:1778-86.

37. Moyer VA, on behalf of the U.S. Preventive Services Task Force. Screening for prostate cancer: U.S. Preventive Services Task Force recommendation statement. *Ann Intern Med* 2012;**157**(2):120-34. doi: 10.7326/0003-4819-157-2-201207170-00459

38. Brito JP, Morris JC, Montori VM. Thyroid cancer: zealous imaging has increased detection and treatment of low risk tumours. *BMJ* 2013;**347**:f4706. doi: 10.1136/bmj.f4706

Table 1. Criteria for assessment of study quality for quantitative studies

	Information Extracted
Study design	Independent-sample or paired-sample design If independent: whether groups were randomised?
Study setting	Community sample, clinics, hospital, other
Selection bias	Independent sample: Were participants allocated to each group randomly? Were samples similar in terms of important characteristics? How was randomisation done? Was randomisation done with concealment allocation? Was the study sample described? Paired-sample: Were terms randomised? How was randomisation done? Was randomisation done with concealment allocation? Was the study sample described?
Performance bias	Were participants kept blind to the study hypothesis?
Attrition bias	Was attrition or exclusions (lost to follow-up) reported?
Reporting bias	Examination of selective reporting. Did they study have a protocol?
Measurement bias	Exposure variable – describe the term used Outcome measures – how were outcomes measured? Were they validated or referenced? Confounders – reporting of additional measures and/or demographics
Analysis	Were appropriate statistical tests used to analyse data and report results?

Table 2. Characteristics of included studies

Study	Year	Country	Disease Focus	Study Design	Study Sample & Setting	Study Aims	Terms Manipulated	Primary Outcomes Measured	Other Outcomes Measured
Copp et al.	2017	Australia	PCOS	Randomised experimental design (2x2x(2) factorial design)	181 female university students	To test the impact of the PCOS disease label on intention to undergo an ultrasound and psychosocial outcomes	'polycystic ovary syndrome' vs 'hormonal imbalance'	Intention to have an ultrasound	Negative affect, self-esteem, perceived severity of condition, credibility of GP, and interest in a second opinion
McCaffery et al.	2015	Australia	DCIS	Randomised experimental design (cross-over design)	269 health women from a community sample	To examine whether the use of terminology including the term cancer to describe DCIS increased hypothetical level of concern and treatment preferences	'abnormal cells' vs 'pre-invasive breast cancer cells'	Treatment preferences: treatment vs watchful waiting, measured on a 5-point Likert scale (definitely prefer treatment, probably prefer treatment, prefer to do nothing, probably prefer watchful waiting, definitely prefer watchful waiting)	Level of concern measured on a 5-point Likert scale (Extremely concerned-Not concerned at all)
Nickel et al.*	2015	Australia	DCIS	Semi-structured qualitative interviews	26 women from a community sample	To understand how different proposed terminologies for DCIS affect women's perceived concern and management preferences	'abnormal cells' vs 'pre-invasive breast cancer cells' 'ductal carcinoma in situ' 'ductal intraepithelial neoplasia' 'indolent lesions of epithelial origin'	Women's qualitative responses to terminologies with and without the cancer term on level of concern and management preferences	
Querner et al.	2013	USA	DCIS	Randomised experimental design (cross-over design)	394 healthy women with no history of breast cancer from a hospital patient registry	To examine how women respond to terminology for DCIS without the cancer term	'non-invasive breast cancer', 'breast lesion', 'abnormal cells'	Treatment preferences (choice between surgery, medication, active surveillance)	
Scherer et al.	2013	USA	GERD	Randomised experimental design (2x2 factorial design)	175 parents aged 18 years presenting at a primary care pediatric clinic	To determine if the disease label GERD influences parents' perceived need to medicate an infant	'GERD' vs no label 'this problem'	Parents' interest in giving their infant medication (3 measures using 6-point Likert scale: No, definitely not-Yes, definitely)	Perception of illness severity (3 measures using 5-point Likert scale: worry = Not at all worried-Very worried, serious = Not at all serious-Very serious, sick = Strongly disagree-Agree)

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49

									Appreciation of medication offered (1 measure using 5-point Likert scale: No, definitely not-Yes, definitely)
Scherer et al.	2015	USA	Conjunctivitis	Randomised experimental design (2x2 factorial design)	159 parents aged 18 years presenting at a primary care pediatric clinic	To determine whether the 'pink eye' term would influence parents' beliefs about the condition and their interest in using antibiotics	'pink eye' vs 'eye infection'	Parents' decision to give their child antibiotics measured on a 6-point Likert scale (No, definitely not-Yes, definitely)	Targeted beliefs about contagiousness measured on a 5-point Likert scale (Not at all contagious-Very contagious) Parents ability to send their child to day care measured on a 5-point Likert scale (Not at all likely-very likely)
Azam et al.	2010	UK	Bony fracture	Cross sectional survey	100 adult patients presenting at an emergency department	To assess the way different terms used to describe a fracture affect the understanding a patients has of it including the perceived severity of the injury and how the patient expects to be treated	'a crack in the bone' 'a broken bone' 'a fracture' 'a hairline fracture' 'a greenstick fracture'	Expected treatment (choice between heals on own, sling, cast, operation)	Perceived severity on a 10-point scale (1 being minimally problematic to 10 being a very serious injury)

*=included qualitative study

Table 3. Risk of bias summary for quantitative studies

Study	Study design	Study setting	Selection bias	Performance bias	Attrition bias	Reporting bias	Measurement Bias	Analysis
Copp, 2017	+	+	+	?	+	+	+	+
McCaffery, 2015	+	+	+	?	+	+	+	+
Omer, 2013	+	?	+	?	?	+	+	+
Scherer, 2013	+	+	+	?	?	+	+	+
Scherer, 2015	+	+	+	?	?	+	+	+
Azam, 2010	-	+	-	?	?	-	-	-

Table 4. Key findings from individual studies

Study	Primary Outcome	Primary Outcome by Terminology		Primary Outcome Statistical Significance*	Other Outcomes
		Medicalised term	Non-medicalised term		
Copp, 2017 (n=181)	Intention to have an ultrasound	Polycystic ovary syndrome (n=90) mean=6.62 (on a 10-point Likert scale where 1=Definitely will not to 10=Definitely will)	Hormonal imbalance (n=91) mean=5.76 (on a 10-point Likert scale where 1=Definitely will not to 10=Definitely will)	Primary outcome significant (p=0.033). Women who received 'PCOS' term in the scenario had significantly higher intentions to have an ultrasound than those who received the 'hormonal imbalance' term F(1,176)=4.63, p=0.033. After women received information on overdiagnosis both intention and perceived severity decreased, regardless of the terminology of the condition (both p<0.001).	Self-esteem: Women's self-esteem was significantly lower for those given the term 'PCOS' than those given the term 'hormonal imbalance' F(1,176)=4.74, p=0.031. Perception of severity: Women who were given the term 'PCOS' had significantly higher perceived severity about the condition than those given the term 'hormonal imbalance' F(1,176)=5.64, p=0.019. Negative affect, credibility of doctor and interest in a second opinion: No difference between terms (all p>0.05).
McCaffery, 2015 (n=269)	Treatment preference	Pre-invasive breast cancer cells (n=128) 40% (51) prefer treatment 60% (77) prefer watchful waiting Change in terminology 41% (52) prefer treatment 59% (76) prefer watchful waiting	Abnormal cells (n=141) 33% (47) prefer treatment 67% (94) prefer watchful waiting Change in terminology 45% (63) prefer treatment 55% (78) prefer watchful waiting	Primary outcomes not significant however change in terminology shows a significant difference. There were no significant difference in treatment between arm A (women who were given the term 'abnormal cells' first and then were given the term 'pre-invasive cancer cells') and arm B (women who were given the term 'pre-invasive cancer cells' first and then were given the term 'abnormal cells') with 33% and 41% of women respectively favouring treatment, p=0.23. In arm A, 18% of women changed their preference to treatment while only 6% changed to watchful waiting (p=0.008). No significant treatment preferences were observed in arm B (9% vs 8%, p>0.99).	Level of concern: There was no significant difference between arms with 49% and 44% of women indicating they would be extremely concerned p=0.600. However, when the alternative term was used women in arm A ('abnormal cells' terminology first and then 'pre-invasive cancer cells' terminology) were significantly more likely to report increased concern than women in arm B ('pre-invasive cancer cells' terminology first and then 'abnormal cells' terminology) 67% vs 52%, p=0.001.
Her, 2013	Treatment preference	Cancer 47% (186) surgery	Lesion 34% (136) surgery	Primary outcome significant (p=<.001)	

1
2
3
4

5 6 7 8 9 10 11		20% (79) medication 33% (129) active surveillance	18% (70) medication 48% (188) active surveillance Abnormal cells 31% (124) surgery 21% (82) medication 48% (188) active surveillance	When DCIS was described as a 'non-invasive cancer' 53% (208 of 394) of participants preferred non-surgical options, whereas 66% (258 of 394) chose non-surgical options when it was described as 'breast lesion' and 69% (270 of 394) chose non-surgical options when it was described as 'abnormal cells' (p<.001).	
12 13 14 15 16 17 18 19 20 21 22	Parents' interest in medicating their infant (3 items pooled and mean reported)	GERD term (n=87) mean=2.51 (on a 6-point Likert scale where 0=No, definitely not to 5=Yes, definitely)	No label (n=88) mean = 2.04 (on a 6-point Likert scale where 0=No, definitely not to 5=Yes, definitely)	Primary outcome significant (p<.01) Parents who received the 'GERD' term in the scenario were more interested in medication than parents who did not receive that term, F(1,165)=6.95, p<.01. Parents not given the 'GERD' term were interested in medication only when they were not given information about medication effectiveness therefore they were allowed to assume that the medications are effective F(1,165)=4.52, P<.05.	Perception of illness severity: Findings were not influenced by the presence or absence of the 'GERD' term (all p>.12). Appreciation of medication offer: Parents were least appreciative of medication when they were told that there medication was ineffective and were not given the 'GERD' term F(1,165)=7.16, p<.01
23 24 25 26 27 28 29 30 31 32 33 34 35	Parents' decision to give their child antibiotics	Pink-eye (n=82) mean=2.43 (on a 6-point Likert scale where 0=No, definitely not to 5=Yes, definitely)	Eye infection (n=76) mean=2.32 (on a 6-point Likert scale where 0=No, definitely not to 5=Yes, definitely)	Primary outcome not significant however when adjusted for information of medication effectiveness outcomes are significant. Parents were less willing to give their child antibiotics when symptoms were referred to as an 'eye infection' information about antibiotic ineffectiveness significantly reduced F(1, 62)=14.67, p<.001. By contrast when parents were told that the symptoms were 'pink eye' interest in antibiotics was not reduced by information about antibiotic ineffectiveness F(1,74)=0.93 p=.33.	Beliefs about contagiousness: Parents who received the 'pink-eye' term thought that the symptoms were significantly more contagious than parents who received the 'eye-infection' term, F(1, 137)=11.21, p=.001. Feelings towards sending child to day care: Parents who received the 'pink-eye' term were less likely to believe that their child would be allowed to go to child care than parents who received the 'eye-infection' term, F(1, 141)=9.70, p=.002.
36 37 38 39 40	Expected treatment	Broken bone 3% heals on own 39% sling 56% cast 2% operation	Crack in the bone 24% heals on own 57% sling 13% cast 6% operation	t-tests which demonstrate significance: Crack vs break=p<0.0001 Crack vs hairline fracture=p<0.0001 Crack vs greenstick fracture=p<0.0001 Break vs hairline fracture=p=0.0001	Patients perspectives of severity: (mean/median score, no SDs given): 3.28/3 for 'a crack in the bone' 3.58/3 for 'a hairline fracture' 4.95/5 for 'a fracture'

41
42
43
44
45
46
47
48
49

1
2
3
4

5					
6		Greenstick fracture	Hairline fracture	Break vs greenstick fracture= $p<0.0001$	5.28/5 for 'a greenstick fracture'
7		47% heals on own	45% heals on own	Fracture vs hairline fracture= $p<0.0001$	6.64/7 for 'a broken bone'
8		25% sling	29% sling	Fracture vs greenstick fracture= $p<0.0001$	
9		22% cast	26% cast		
10		6% operation	0% operation		
11					
12		Fracture			
13		14% heals on own			
14		44% sling			
15		36% cast			
16		6% operation			

17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49

*as reported by original study authors
**=1 case from total sample missing

For peer review only

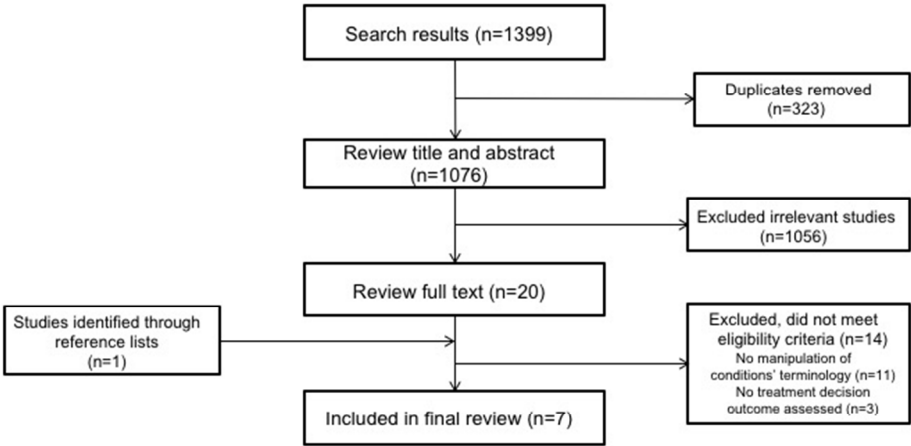
Table 5. Summary of preferences for more invasive management option by type of terminology*

Study	More medicalised or more precise term/s (%)	Less medicalised or less precise term/s (%)	Difference (%)	p-value
Copp, 2017	70	53	17	>0.05§
McCaffery, 2015	40	33	7	0.23
Omer, 2013	47	32.5	14.5	<0.001
Scherer, 2013	74	67	7	>0.1§
Scherer, 2015	60	58	8	>0.1§
Azam, 2010	39 (4 operation, 35 cast)†	19 (6 operation, 13 cast)‡	20	<0.025§

*=combined data where applicable and mean percentages reported, see Appendix 4 for explicit justification of categorisation of terminology
†=broken bone, fracture, greenstick fracture, hairline fracture; ‡=crack in the bone
§=calculated significance using raw (Copp and Scherer) and published (Azam) data based on our classification of which terms were more medicalised

Figure 1. Flow diagram of study selection

For peer review only



254x190mm (72 x 72 DPI)

Appendix 1. Search strategy

OVID Medline	OVID Pre-Medline	Embase	PsycINFO	CINAHL	PubMed
1. exp Terminology as Topic/ 2. terminolog*.tw. 3. medical term*.tw. 4. disease label*.tw. 5. (disease adj3 label*).tw. 6. medical label*.tw. 7. 1 or 2 or 3 or 4 or 5 or 6 8. Patient Preference/ 9. Decision Making/ 10. patient decision making.tw. 11. (patient adj3 decision*).tw. 12. treatment decision making*.tw. 13. (treat* adj3 decision*).tw. 14. treatment pref*.tw. 15. (treat* adj3 pref*).tw. 16. management pref*.tw. 17. (manage* adj3 pref*).tw. 18. overdiagnosis.kw. 19. 20. overtreatment.kw. 21. 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15 or 16 or 17 or 18 or 19 22. 7 and 20	1. terminolog*.tw. 2. medical term*.tw. 3. disease label*.tw. 4. (disease adj3 label*).tw. 5. medical label*.tw. 6. 1 or 2 or 3 or 4 or 5 7. patient preference/ 8. patient decision making.tw. 9. treatment decision making*.tw. 10. (treat* adj3 decision*).tw. 11. treatment pref*.tw. 12. (treat* adj3 pref*).tw. 13. management pref*.tw. 14. (manage* adj3 pref*).tw. 15. overdiagnosis.kw. 16. overtreatment.kw. 17. 7 or 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15 or 16 18. 6 and 17	1. nomenclature/ 2. terminolog*.tw. 3. medical term*.tw. 4. disease label*.tw. 5. (disease adj3 label*).tw. 6. medical label*.tw. 7. 1 or 2 or 3 or 4 or 5 or 6 8. patient preference/ 9. patient decision making/ 10. treatment decision making*.tw. 11. (treat* adj3 decision*).tw. 12. treatment pref*.tw. 13. (treat* adj3 pref*).tw. 14. management pref*.tw. 15. (manage* adj3 pref*).tw. 16. 17. overtreatment.kw. 18. 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15 or 16 or 17 19. 7 and 18	1. terminology/ 2. terminolog*.tw. 3. medical term*.tw. 4. disease label*.tw. 5. (disease adj3 label*).tw. 6. medical label*.tw. 7. 1 or 2 or 3 or 4 or 5 or 6 8. decision making/ or management decision making/ 9. patient decision making.tw. 10. treatment decision making*.tw. 11. (treat* adj3 decision*).tw. 12. treatment pref*.tw. 13. (treat* adj3 pref*).tw. 14. management pref*.tw. 15. (manage* adj3 pref*).tw. 16. overdiagnosis.kw. 17. 18. overtreatment.kw. 19. 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15 or 16 or 17 20. 7 and 18	S1 TI terminolog* or AB terminolog* S2 TI medical term* or AB medical term* S3 TI disease label* or AB disease label* S4 TI disease N3 label* or AB disease N3 label* S5 TI medical label* or AB medical label* S6 S1 OR S2 OR S3 OR S4 OR S5 S7 (MH "Decision Making") OR (MH "Decision Making, Patient") S8 TI treatment decision making or AB treatment decision making S9 TI treatment N3 decision* or AB treatment N3 decision* S10 TI treatment pref* or AB treatment pref* S11 TI treatment N3 pref* or AB treatment N3 pref* S12 TI management pref* or AB management pref* S13 TI manage* N3 pref* or AB manage* N3 pref* S14 "overdiagnosis" S15 "overtreatment" S16 S7 OR S8 OR S9 OR S10 OR S11 OR S12 OR S13 OR S14 OR S15 S17 S6 AND S16	((((terminolog*[Title/Abstract]) OR medical term*[Title/Abstract]) OR medical label*[Title/Abstract]) OR disease label*)) AND ((((treatment decision making[Title/Abstract]) OR treatment decision*[Title/Abstract]) OR treatment pref*[Title/Abstract]) OR management pref*[Title/Abstract]) OR overdiagnosis) OR overtreatment)

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Appendix 2. Eligibility criteria (with exclusion criteria)

Types of studies:	Empirical studies that considered management decision making as an anticipated or experienced outcome of being given or told one terminology vs another for the same condition including surveys, questionnaires and interviews or focus groups <i>Exclusion: review papers, editorials, commentary/discussion papers</i>
Types of participants/population:	Adults (patients, community members) aged 18 years and above including adults making surrogate decisions or decisions on behalf of their children <i>Exclusion: patients or community members less than 18 years of age making a decision for themselves</i>
Types of settings:	Any type of medical or community setting including hypothetical scenarios given to community members
Study factor (intervention):	Different term given for the same condition
Outcome factor:	Treatment or management preferences (e.g. treatment 1 vs treatment 2 vs treatment 3/no treatment)

Appendix 3. Nickel et al. qualitative study appraisal*

Were the researcher characteristics described?	Partly/moderate quality – interviews were conducted by two public health researchers with experience in qualitative research methods, however credentials, gender and characteristics of the interviewer were not reported
Was the methodological orientation to the study described?	Yes/ high quality – rationale informed by previous published DCIS terminology research (Esserman 2013, Omer 2013)
Was the target population clearly defined?	Yes/high quality – community sample of 26 Australian women aged 25 years and above who were recruited from a national community survey. Table 1 includes participant's characteristics (age, education, experience with cancer screening, previous breast cancer diagnosis, immediate family history of cancer, employment status, worked as a health professional, survey management preference)
Was the sampling strategy clearly defined?	Yes/high quality – consenting women were recruited from a national community survey and purposively selected according to their education background (lower vs higher education), previous screening experience (previously screened vs not) and management preferences (immediate treatment vs watchful waiting as indicated by their survey responses)
Were the data collection methods clearly described?	Yes/high quality – semi-structured single telephone interview which lasted 13-40 minutes, and were audio-recorded and transcribed verbatim
Was data analysis clearly described?	Yes/high quality – Framework Analysis was used to organise the data and identify main themes that capture the diverse views expressed with two researches independently coding data using a rigorous analysis process

*criteria adapted from Tong et al.¹²

Appendix 4. Justification for management and terminology classification in each study

Study	Management	Classification	Justification	Terminology	Classification	Justification
Copp, 2017*	6-10 on Likert scale "Intention to have a screening test (ultrasound)"	Interest in medical management	Likert scale with anchored end-points (1=Definitely will not; 10=Definitely will), assumed 1-5=No intention, 6-10=Intention	Polycystic ovary syndrome	More medicalised term	The condition's full medical terminology (PCOS) was deemed to be more medicalised.
	1-5 on Likert scale "Intention to have a screening test (ultrasound)"	No interest in medical management		Hormonal imbalance	Less medicalised/less precise term	
McCaffery, 2015*	Treatment	Invasive treatment		Pre-invasive breast cancer cells	More medicalised term	Terminology which included the cancer term was deemed to be more medicalised than without.
	Watchful waiting	Non-invasive treatment		Abnormal cells	Less/non-medicalised term	
Omer, 2013*	Surgery	Invasive treatment		Cancer	More medicalised term	Terminology which included the term cancer was deemed to be more medicalised than those without.
	Medication Active surveillance	Non-invasive treatment		Lesion Abnormal cells	Less/non-medicalised term	
Scherer, 2013*	2-5 on Likert scale "Will you give your infant this medication?"	Invasive treatment	Likert scale with anchored end-points (0=No, definitely not; 5=Yes, definitely), assumed 1=No, 2=Maybe not, 3=Maybe, 4=Yes†	Gastroesophageal reflux disease (GERD)	More medicalised term	The condition's full medical terminology (GERD) was deemed to be more medicalised.
	0+1 on Likert scale "Will you give your infant this medication?"	Non-invasive treatment		"This condition" (no label)	Less/non-medicalised term	
Scherer, 2015*	2-5 on Likert scale "Will you give your child antibiotics?"	Invasive treatment	Likert scale with anchored end-points (0=No, definitely not; 5=Yes, definitely), assumed 1=No, 2=Maybe not, 3=Maybe, 4=Yes†	Pink-eye	More precise medical term	Giving the condition a specific terminology (pink-eye) rather than a generic term (eye infection) was deemed to be more medicalised.
	0+1 on Likert scale "Will you give your child antibiotics?"	Non-invasive treatment		Eye infection	Less precise term	

Azam, 2010	Operation Cast	Invasive treatment		Broken bone Fracture Greenstick fracture Hairline fracture	More medicalised term	Authors determined classification based on information on the precise terminology healthcare professionals use to describe a bony injury.
	Heals on own Sling	Non-invasive treatment		Crack in the bone	Less/non- medicalised term	

*=studies where authors were specifically guided by study design to classify terminology

†=sensitivity analysis conducted



PRISMA 2009 Checklist

Section/topic	#	Checklist item	Reported on page #
TITLE			
Title	1	Identify the report as a systematic review, meta-analysis, or both.	1
ABSTRACT			
Structured summary	2	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.	2-3
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known.	5
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).	6
METHODS			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.	6
Eligibility criteria	6	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	7 (Appx.2)
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	6-7
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	6-7 (Appx.1)
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	7
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	6-7
Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.	6-7
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.	7 (Table 1 & Appx. 3)
Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means).	8
Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., I^2 for each meta-analysis)	8 (Appx. 4)



PRISMA 2009 Checklist

Page 1 of 2

Section/topic	#	Checklist item	Reported on page #
Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).	N/A
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.	
RESULTS			
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.	8
Study characteristics	18	For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.	8 (Table 2)
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12).	8-9 (Table 3)
Results of individual studies	20	For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot.	9-14 (Table 4)
Synthesis of results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.	14 (Table 5)
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see Item 15).	N/A
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]).	
DISCUSSION			
Summary of evidence	24	Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy makers).	14-15
Limitations	25	Discuss limitations at study and outcome level (e.g., risk of bias), and at review-level (e.g., incomplete retrieval of identified research, reporting bias).	16-17
Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.	18
FUNDING			
Funding	27	Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review.	19

From: Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med 6(7): e1000097. doi:10.1371/journal.pmed1000097

For peer review only - <http://bmjopen.bmj.com/site/about/guidelines.xhtml>

For more information, visit: www.prisma-statement.org. Page 2 of 2

BMJ Open

Words do matter: a systematic review on how different terminology for the same condition influences management preferences

Journal:	<i>BMJ Open</i>
Manuscript ID	bmjopen-2016-014129.R2
Article Type:	Research
Date Submitted by the Author:	20-May-2017
Complete List of Authors:	Nickel, Brooke; University of Sydney, School of Public Health Barratt, Alexandra; University of Sydney, School of Public Health Copp, Tessa; University of Sydney School of Public Health, Moynihan, Ray; Bond University , Centre for Research in Evidence-Based Practice McCaffery, Kirsten; The University of Sydney, Screening and Test Evaluation Program (STEP), School of Public Health; The University of Sydney, Centre for Medical Psychology & Evidence-based Decision-making (CeMPED)
Primary Subject Heading:	Communication
Secondary Subject Heading:	Medical management
Keywords:	terminology, management preferences, treatment preferences, decision making, overtreatment

SCHOLARONE™
Manuscripts

only

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Words do matter: a systematic review on how different terminology for the same condition influences management preferences

Brooke Nickel *PhD candidate*¹, Alexandra Barratt *epidemiologist*¹, Tessa Copp *PhD candidate*¹, Ray Moynihan *research fellow*^{1,2}, Kirsten McCaffery *health psychologist and professorial research fellow*¹

¹Wiser Healthcare, Sydney School of Public Health, The University of Sydney, Sydney, New South Wales, Australia

²Centre for Research in Evidence-Based Practice, Bond University, QLD 4229, Australia

Correspondence to: Ms Brooke Nickel; Wiser Healthcare, Sydney School of Public Health, The University of Sydney, NSW 2006, Australia. (email: brooke.nickel@sydney.edu.au)

ABSTRACT

Objectives: Changing terminology for low-risk, screen detected conditions has now been recommended by several expert groups in order to prevent overdiagnosis and reduce the associated harms of overtreatment. However, the effect of terminology on patients' preferences for management is not well understood. This review aims to synthesise existing studies on terminology and its impact on management decision making.

Design: Systematic review.

Methods: Studies were included that compared two or more terminologies to describe the same condition and measured the effect on treatment or management preferences and/or choices. Studies were identified via database searches from inception to April 2017, and from reference lists. Two authors evaluated the eligibility of studies with verification from the study team, extracted and cross-checked data, and assessed the risk of bias of included studies.

Results: Of the 1399 titles identified, 7 studies, all of which included hypothetical scenarios, met the inclusion criteria. Six studies were quantitative and one was qualitative. Six of the studies were of high quality. Studies covered a diverse range of conditions: ductal carcinoma in situ (3), gastroesophageal reflux disease (1), conjunctivitis (1), polycystic ovary syndrome (1) and a bony fracture (1). The terminologies compared in each study varied based on the condition assessed. Based on a narrative synthesis of the data, when a more medicalised or precise term was used to describe the condition it generally resulted in a shift in preference towards more invasive managements, and/or higher ratings of anxiety and perceived severity of the condition.

Conclusions: Different terminology given for the same condition influenced management preferences and psychological outcomes in a consistent pattern in these studies. Changing the terminology may be one strategy to reduce patient preferences for aggressive management responses to low-risk conditions.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Registration: CRD42016035643.

For peer review only

STRENGTHS AND LIMITATIONS OF THIS STUDY:

- This is the first systematic review to synthesise the evidence on how different terminology given for the same condition impacts patients' management preferences
- Only a small number of studies have examined this research question and were included in the review
- Due to the variability of terms and outcomes assessed, authors were unable to conduct a meta-analysis and pool the effects of the data
- All studies included were hypothetical, therefore patients facing a real diagnosis may respond differently

INTRODUCTION

Medical encounters can be challenging and confronting for patients, especially when they are faced with a management decision. Clinical communication and language is an important aspect of a medical encounter as it influences patients’ understanding of their diagnosis and management options¹². Decisions about treatments or tests may be influenced by various communication factors including the medical terminology clinicians use to diagnose and describe conditions to patients.

Overdiagnosis of several medical conditions and associated overtreatment is now widely accepted^{3 4}, and can have serious implications for patients, healthcare systems and society^{5 6}. Numerous approaches are beginning to be proposed to help combat overdiagnosis and overtreatment, including various communication strategies⁷. Changing the terminology for medical conditions may be one communication strategy to mitigate the effect of overdiagnosis and overtreatment as it has the potential to influence the effect of diagnosis and labelling of a condition, and influence patients’ decision making about management. It may encourage both patients and clinicians to more carefully consider conservative management options.

In particular, cancer terminology is one area where use of different terminology may greatly influence management decision making. The term ‘cancer’ is understandably frightening for people to hear and can influence their thought and action⁸, but it is now well accepted by cancer experts, researchers and clinicians that a range of conditions which include indolent to fast-growing lesions are labelled as cancer⁹. Lesions with low malignant potential are common (such as low-risk DCIS, low-risk papillary thyroid cancer and low-grade prostate cancer), and with the advent and increasing use of various screening technologies these indolent lesions and their precursors are now frequently clinically identified. This identification can lead to the condition being overdiagnosed and in turn overtreated.

Changing the terminology for these low-risk, screen detected conditions has now been recommended by several expert groups – including a National Institutes of Health state of the science conference panel and a National Cancer Institute working group – in order to prevent overdiagnosis and encourage more consideration of less invasive management options such as active surveillance⁹⁻¹¹. However, the effect of terminology on patients' willingness to accept more conservative management options is not well understood. This systematic review aims to synthesise existing studies on terminology for medical conditions and its impact on management decision making, and associated psychosocial outcomes.

METHODS

Protocol and registration

The review's protocol is registered with PROSPERO (an international prospective register of systematic reviews), registration number: CRD42016035643.

Review Question

How do different terminologies given to the same condition influence treatment or management decision making?

Search Strategy

A comprehensive list of search terms was developed (Appendix 1) with consultation from an information specialist and a search of relevant databases (Medline, Pre-Medline, Embase, PsycINFO, Cinhal, and PubMed) was conducted from inception to April 2017. The returned search results were screened by title and abstract independently by two researchers (BN and TC) for irrelevant articles, reviews papers, editorials or commentaries, and duplicates. An eligibility checklist was developed (Appendix 2) to guide the selection of appropriate studies. Decisions regarding inclusion and exclusion of studies was then made independently by two researchers (BN and TC) and

disagreements discussed. Any further disagreement or uncertainty was discussed and verified by two additional researchers (AB and KM). A hand search of reference lists of included studies as well as papers recommended through personal communication were also examined for relevant studies.

Inclusion and Exclusion Criteria

Studies were included in the review if they empirically measured treatment or management decision making as an anticipated or experienced outcome of being given or told one terminology versus another or others for a specific condition. Studies were excluded if they were reviews, editorials or commentaries, or if they assessed participants less than 18 years of age making a decision for themselves (Appendix 2).

Quality Assessment and Data Extraction

All studies that met the inclusion criteria were appraised for study quality by two authors (BN and TC) independently using two separate tools; one for quantitative studies and one for the qualitative study. The quantitative studies were appraised using a modified version of the Cochrane Collaboration’s tool for assessing risk of bias which was developed by study authors (Table 1). As the quantitative studies included in this review were not clinical trials, study authors adapted the Cochrane Collaboration’s tool for assessing risk of bias to appraise the studies for items relating to study design, study setting, study validity, and analysis. Study validity included allocation of participants to different terminology, blinding of participants to the study hypotheses, and reporting and measurement bias. The qualitative study was appraised using criteria adapted from the Consolidated criteria for reporting qualitative studies (COREQ) framework¹² and results appear in Appendix 3.

Two categories of study quality were identified by study authors according to each study’s methodological characteristics. In high-quality studies (lower risk of bias) the majority of criteria

1
2
3 were fulfilled and done well, while in low-quality studies (higher risk of bias) the majority of criteria
4
5 were not done or done poorly.
6
7

8 Data from the final studies included in the review were extracted independently into a standardised
9
10 template by two authors (BN and TC). Any discrepancies were discussed and resolved by the entire
11
12 research team. For studies where the data was not reported or was unclear, authors were contacted
13
14 and responded for confirmation of details of the data. Authors were also asked to confirm the
15
16 extracted data to minimise any reporting bias. All authors responded. Results from the studies were
17
18 synthesised in a narrative form, as the heterogeneity of the studies and their respective outcome
19
20 measures did not support pooling of results¹³.
21
22

23 24 RESULTS

25
26 Initial search results identified 1399 papers. After removal of duplicates and screening by title and
27
28 abstract, 20 studies from the search and 1 study identified from reference lists underwent full-text
29
30 reviews. Seven studies that reported the impact of different terminology for the same condition on
31
32 treatment or management decision making were identified (Figure 1). Three studies reported on
33
34 ductal carcinoma in situ (DCIS)¹⁴⁻¹⁶, two on common childhood conditions (gastroesophageal reflux
35
36 disease (GERD), conjunctivitis)¹⁷⁻¹⁸, one on polycystic ovary syndrome (PCOS)¹⁹ and one on a bony
37
38 fracture²⁰. Six studies reported quantitative findings^{14 15 17-20} and one study reported qualitative
39
40 findings¹⁶. The key characteristics of these studies and their methods are summarised in Table 2. All
41
42 studies were hypothetical and involved various samples of community members who were not
43
44 currently and/or previous diagnosed with the condition assessed. Of the six quantitative studies, five
45
46 involved a randomised experimental design; with two using a paired sample design (cross-over) and
47
48 three using an independent sample design (2x2 factorial). The qualitative study was linked to one of
49
50 the quantitative studies¹⁵, as women from the study were asked if they would be willing to
51
52 participate in an additional qualitative interview for further investigation of the topic.
53
54
55
56
57
58
59
60

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Overall the majority of studies were of higher quality with a lower risk of bias, however one study did have a high risk of bias as the quality of reporting was low²⁰ (Table 3). Higher quality studies had defined study populations and settings, had low selection bias and described the data collection, methods and analysis appropriately. The qualitative study included in the review was rigorous and had appropriate methods to reach its research objectives. Researchers and the target population were defined, the sampling strategy was explained, data collection methods and data analysis were appropriate and well documented.

Since the strength of the review is the diversity of included studies, results of the studies are summarised individually by study quality category, with data on management outcomes summarised in Table 4. Following this is a narrative synthesis of the effect of terminology (more medicalised or precise terms vs less or non-medicalised terms) on management preferences (invasive management preference vs non-invasive management preference). Importantly for each study we identify the classifications of management preferences and terminology and provide justification where applicable (Appendix 4).

Results from individual higher quality studies

Copp et al. 2017 (Polycystic ovary syndrome)

This study on the influence of the term PCOS found that when young women were given the term ‘PCOS’ in a hypothetical scenario of a doctor’s visit, they had significantly higher intention to have an ultrasound compared to women who were given the term ‘hormonal imbalance’ (mean=6.62 vs. mean=5.76, $F(1, 176)=4.63, p=0.033$). The study also found that those who received the ‘PCOS’ term perceived their hypothetical condition to be more severe (17.7 vs. 15.82, $F(1, 176)=5.65, p=0.019$) and had lower self-esteem compared to women who were not given the PCOS term (25.86 vs. 27.56, $F(1, 176)=4.74, p=0.031$). After women received information about the potential of PCOS

overdiagnosis in a second scenario, both intention and perceived severity decreased, regardless of the term given (both $p < 0.001$).

The study also found a significant 3-way interaction between the term 'PCOS', information about ultrasound reliability and overdiagnosis information ($F(1, 176) = 4.23$, $p = 0.041$), where that for those who did not receive the PCOS term, intention was significantly lower for those who received information about the unreliability of ultrasounds compared with those who received no information. For women who received the PCOS term however, intention was high, even when told the ultrasound was unreliable. This difference disappears after information about overdiagnosis is given in the second scenario, suggesting the provision of overdiagnosis information reduces the effect of the term.

McCaffery et al. 2015 (Ductal carcinoma in situ)

McCaffery and colleagues' study on the impact of DCIS terminology on treatment preference (immediate treatment vs watchful waiting) and women's level of concern found no significant differences in treatment preference between arm A (women who were given the term 'abnormal cells' first and then were given the term 'pre-invasive cancer cells') and arm B (women who were given the term 'pre-invasive cancer cells' first and then were given the term 'abnormal cells') with 33% and 40% of women respectively favouring treatment ($p = 0.23$). However, 18% of women in arm A who were initially given the 'abnormal cells' terminology changed their preference to treatment when the terminology was switched to 'pre-invasive breast cancer cells' while only 6% changed to watchful waiting ($p = 0.008$). In contrast, there were no significant changes in treatment preference in arm B when the terminology was switched the other way (9% vs 8% changed their stated preference).

Similarly, this study found that there was no significant difference between arms with regards to level of concern with 49% and 44% of women indicating they would be extremely concerned

p=0.600. However, when the alternative term was used, women in arm A (who were initially given the ‘abnormal cells’ term and then were given the ‘pre-invasive breast cancer cells’ term) were significantly more likely to report increased concern than women in arm B (67% vs 52%, p=0.001). Findings from this study were also supported by the included qualitative study conducted with a subset of women¹⁶.

Nickel et al. 2015 (Ductal carcinoma in situ)

This qualitative study investigated in-depth how different proposed terminologies to describe DCIS affected women’s treatment preferences and psychological outcomes by conducting semi-structured interviews with women of varying education, cancer screening experience and with no history of a DCIS diagnosis. Findings demonstrated that overall women preferred a diagnosis of DCIS to be communicated using terminology that did not include the term cancer, as women generally exhibited stronger negative reactions when the cancer term was used to describe DCIS compared to when a non-cancer term was used. Although concern seemed to be high overall, women displayed a high level of interest in watchful waiting when it was described to them in a hypothetical scenario as a safe and effective option, and told that they could proceed to treatment in the future if necessary.

Omer et al. 2013 (Ductal carcinoma in situ)

This study also examined the impact of DCIS terminology on women’s treatment preferences by comparing 3 identical scenarios, with the only difference being the term used to described DCIS (‘non-invasive cancer’, ‘breast lesion’ and ‘abnormal cells’). Treatment options presented were surgery, medication, active surveillance. All participants saw all scenarios, with the order of scenarios varied across participants. When DCIS was described as a non-invasive cancer 53% of participants preferred non-surgical options, whereas 66% chose non-surgical options when it was described as breast lesion and 69% chose non-surgical options when it was described as abnormal cells (p<0.001). Although women with a previous history of cancer (other than breast cancer) and

women with high socioeconomic status more frequently chose surgery in univariate analyses, high numeracy was the only independent predictor of preference for surgical treatment in the multiple variable logistic regression model for all three terms: cancer (OR 2.11, 1.34-3.34 CI, $p=0.001$), lesion (OR 1.96, 1.20-3.19, $p=0.001$), abnormal cells (OR 1.63, 1.01-2.67, $p=0.048$).

Scherer et al. 2013 (Gastroesophageal reflux disease)

This study on the influence of the term 'GERD' (versus no term or label given) on parents preferences for medication for their infant found that parents who received the term GERD in the scenario were more interested in medication than parents who did not receive that term, ($F(1, 165)=6.95$, $p<0.01$). To assess parental interest in antibiotics, study authors combined three highly inter-correlated questions involving parent's interest in medication: Will you give your infant this medicine? Do you think your infant needs the medicine your doctor offered? Do you think that the medicine will help your infant get better?

The study also found a significant interaction between the term GERD and report of medicine ineffectiveness, ($F(1, 165)=4.52$, $p<0.05$) as parents who received a GERD diagnosis were interested in medicating the infant, even if they were told that the medications are likely ineffective. By contrast parents not given a diagnosis were interested in medication only when they were not given information about medication effectiveness, thus able to assume that the medications are effective.

All parents in the study were also asked (using a 0-5 Likert scale with labelled end-points for each question) whether they were worried about their infant's health (mean=2.28, SD=1.30), thought the condition was somewhat serious (mean=2.12, SD=1.19), and were relatively unlikely to describe their infant as being sick (mean=1.87, SD=1.45). These answers were not influenced however by the presence or absence of the GERD term (all $p>0.12$).

Scherer et al. 2015 (Conjunctivitis)

A similar study design conducted by Scherer and colleagues did not find an initial difference between the term 'pink-eye' and 'eye-infection' on parents' preference to medicate their infant. However, when symptoms were referred to as an 'eye infection', information about antibiotic ineffectiveness significantly reduced interest in using medication ($F(1, 62) = 14.67, p < 0.001$). By contrast, when parents were told that the symptoms were 'pink eye', interest in antibiotics was not reduced by information about antibiotic ineffectiveness ($F(1, 74) = 0.93, p = 0.33$).

Study authors also measured (using a 0-5 Likert scale with labelled end-points) parents' perceptions about contagiousness and belief that their child could attend child care and found that parents who received the 'pink-eye' term thought that the symptoms were significantly more contagious (mean=4.04, SD=1.30) than parents who received the 'eye-infection' term (mean=3.29, SD=1.63), ($F(1, 137) = 11.21, p = 0.001$). Parents who received the 'pink-eye' term were also less likely to believe that their child would be allowed to go to child care (mean=0.35, SD=0.90) than parents who received the 'eye-infection' term (mean=0.93, SD=1.13), ($F(1, 141) = 9.70, p = .002$).

Individual results from the lower quality study

Azam et al. 2010 (Bony Fracture)

This study found that patients' management expectations and perceptions of bony injuries differ based on the terminology used to describe the injury. Patients expected more invasive managements (operation or cast) when a more medicalised term was used to describe the injury compared to a less invasive treatment (sling or heals on own), with 58% of patients expecting invasive treatments for the term 'broken bone', 42% for 'fracture', 28% for 'greenstick fracture' and 26% for hairline fracture. In contrast, patients given the less medicalised term had a lower preference for invasive treatments, with 19% of patients choosing an invasive treatment for 'crack in the bone'. This study also found that patients perceived the injury to be more severe when a more

medicalised term was used to describe the injury (average mean severity score out of 10; broken bone=6.64, greenstick fracture=5.28, fracture=4.95, hairline fracture=3.58, crack in the bone=3.28).

Synthesis of results

Table 5 summarises preferences for more invasive management option by type of terminology across the studies. Generally, there was a pattern in the same direction showing that when more medical or precise term was used it resulted in a greater preference or interest in more invasive managements, whether this be intention to have an ultrasound for a PCOS diagnosis, surgical treatment for DCIS, increased interest in potentially ineffective medication, or an operation or cast for a bony injury, although these differences did not always reach statistical significance. In contrast, when participants were given a less or non-medical/precise term a higher proportion of participants chose a non-invasive management option. In this analysis medicalised or precise terminology refers to language that describes the condition either using medical terms that healthcare professionals commonly used or that described the condition in a more specific way (when compared to the comparator term). Studies varied as some found a significant effect of terminology on management decision making while others found a significant interactions effect or within person effect and psychological outcome effects (e.g. perceived severity, level of concern).

DISCUSSION

This is the first systematic review which synthesises the evidence on how different terminology given for the same condition impacts management preferences. Overall, the review demonstrates that different terminology used to describe the same condition can influence patient's management preferences. Results indicate that when a more medical or more precise term was used to describe a condition, people tended to have stronger preferences for more invasive management options. Although not all of the studies included in our review had results which were statistically significant in relation to preferences for more invasive managements, at a population level these trends may

represent a clinically important difference. For example, a relatively small increase in the number of people preferring surgery in these studies could translate into significantly more surgeries across a larger population.

The terminology given during a diagnosis to patients is important, as when a healthy individual becomes a patient they can immediately become more vulnerable to the words their clinicians use²¹. Using words that generate fear or anxiety cause patients to have more difficulty making informed decisions and becoming an active participant in their care²². Different terms used to describe the same condition can have a direct influence on how patients understand their diagnosis and how threatening they perceive it to be²³. It has also been shown that the use of interpretive terminology (e.g. including the words positive or negative, or using a metaphor)^{24 25}, the terminology used to describe a treatment choice²⁶, describing a condition with plain language terminology as compared to jargon²⁷ and the severity of the characteristics of the diagnosis²⁸ can have an influence on medical decision making. Furthermore, a recent survey of medical students found that students were more likely to classify condition synonyms as a 'disease' if the term used to describe it was medical²⁹. Together with findings from this review, these studies show that language is a powerful tool that has the potential to influence patients' thoughts and actions.

Our findings are also in line with the results of other research on the effect of labelling a condition in relation to the social implications it may have for the individual, rather than its effect on medical decision making. Research has found that disease labels can result in various emotional, cognitive and physical consequences^{30 31}. Most notably, studies which have examined the effect of hypertension labelling have found that giving the label of hypertension to individuals (compared to giving no label) resulted in increased self-reported illness and absenteeism from work, and a significant increase in blood pressure^{30 32 33}. In the conjunctivitis study by Scherer and colleagues included in this review, the term 'pink-eye' was perceived as being more contagious, and parents were less likely to believe their child could go to childcare, compared with parents who received the

condition described as an 'eye-infection'. Additionally, in the PCOS study women had significantly lower levels of self-esteem when the term 'PCOS' was used to describe their condition compared to when it was described as a hormonal imbalance¹⁹.

Using more medicalised or precise medical terminology and disease labels may also reduce a patient's sense of self-control, and therefore increase perceived severity and uptake of medications. The more medical the term or label, the less control a person may feel over the situation, increasing their perceived severity about the condition and creating a perception that more invasive interventions are warranted. This was also shown in the GERD, conjunctivitis and PCOS study where there was a high level of interest in medical intervention when a precise medical term or label was given, even when participants were told that the medication or test was ineffective^{17 18}.

This study was limited by the small number of included studies in the review, which also included three studies that were conducted by members of the review team. However, in light of recent and ongoing evidence of overtreatment across a number of conditions and suggestions from leading global medical bodies to change the terminology of low-risk conditions^{10 11}, this review addressed an important and timely under-researched question.

Changing the terminology of low-risk conditions may be difficult in practice as a systems level approach would need to be taken to ensure that all healthcare professionals implemented the new terminology. Although, it would be feasible as demonstrated by the recent change to the terminology of the 'non-invasive encapsulated follicular variant of papillary thyroid carcinoma' (EFVPTC) to be 'non-invasive follicular thyroid neoplasm with papillary-like nuclear features' (NIFTP) in order to highlight the true nature of the tumour, lessen the emotional and psychological burden associated with the term 'cancer' and potentially reduce overtreatment^{34 35}.

Due to the variability of the terms and outcomes assessed in the included studies, authors were unable to conduct a meta-analysis and pool the effects of the data, and therefore synthesised the

1
2
3 results narratively. Since a quantitative synthesis of the data was not possible, findings may not be
4
5 generalisable. Furthermore, the delineation between more medical/precise and less or non-
6
7 medical/precise was challenging in some studies. For example, it seemed clear that ‘pre-invasive
8
9 breast cancer cells’ was a more medicalised term than ‘abnormal cells’. In contrast, it was not as
10
11 clear what it was exactly about the term ‘pink eye’ (a more precise term to describe the condition)
12
13 that elicited stronger parental preferences for antibiotics when compared with the term ‘eye
14
15 infection’. Other aspects important to parents in this study (e.g. aesthetical aspects), may have
16
17 influenced management preferences. The author team therefore made explicit judgements about
18
19 which terms were deemed more medicalised or precise and which were not, as well as what
20
21 managements were considered invasive and what were considered non-invasive. These decisions
22
23 were guided from the aims and outcomes of the studies, and followed categories used in the original
24
25 studies, with the exception of the Azam paper where authors were guided by information on the
26
27 precise medical terminology healthcare professionals use to describe a bony injury (including
28
29 ‘broken bone’, ‘fracture’, ‘greenstick fracture’, ‘hairline fracture’) (see Appendix 4). We note the
30
31 Azam paper was of lower quality which made interpretations about the terms used more difficult.
32
33 Nonetheless, it was very clear that in all of the studies the use of different terminology for the same
34
35 condition tended to elicit different responses to management preferences as well as psychosocial
36
37 outcomes. Further research could usefully explore in more detail what characterises terms that elicit
38
39 stronger preferences for more invasive managements.
40
41
42
43
44
45 Importantly, all studies included in this review were hypothetical and were not presented to patients
46
47 randomised to receive a diagnosis and management preferences were assessed instead of actual
48
49 management decision making. Patients facing real decisions may respond differently to those in the
50
51 studies, however it is likely that in real life situations these effects may be even more pronounced. It
52
53 is likely that patients would be more anxious than participants in current studies, therefore more
54
55 susceptible to judgement biases that result from more medicalised terminology or labels.
56
57
58
59
60

1
2
3 Against a backdrop of recent evidence of overdiagnosis and overtreatment across a number of
4
5 conditions^{5 36-38} and the potential physical and psychological impacts this may have for the patient,
6
7 understanding how different terminology given for the same medical condition may influence
8
9 patients' management preferences is important. This review suggests that the terminology used to
10
11 describe a condition can influence patient preferences for management and related outcomes.
12
13 Although further evidence is needed to help better understand precisely why some terms elicited
14
15 stronger preferences for more invasive management, this review helps support the calls for changing
16
17 the terminology of conditions where the risk of progression is low. Changing the terminology in low-
18
19 risk conditions or conditions with indolent clinical course may be a potential communication strategy
20
21 to help shift assumptions that immediate invasive treatments or tests are always needed, allow for
22
23 better shared decision making between clinicians and patients, and the consideration of more
24
25 conservative management options.
26
27
28
29
30
31
32

33 **Acknowledgements:** The authors would like to thank Philippa Bourke for her help with the
34
35 literature search and Dr Kevin McGeechan for his comments on the statistical aspects of the paper.
36
37

38 **Contributions:** BN, AB and KM contributed to study concept and design. BN conducted the
39
40 systematic literature search. BN and TC evaluated the eligibility of studies, performed the study
41
42 quality assessment, extracted the data and conducted the evidence synthesis. BN, AB, TC, RM and
43
44 KM interpreted the findings. BN drafted the manuscript and all authors critically reviewed and
45
46 approved the manuscript.
47
48

49 **Funding:** This paper was written with support from the Wiser Healthcare CRE grant from the
50
51 National Health and Medical Research Council (NHMRC) #1104136. BN was supported by the Sydney
52
53 Catalyst Research Scholar Award. KM was supported by an NHMRC fellowship (1029241). The
54
55
56
57
58
59
60

fundere had no role in the design or conduct of the study; in the collection, analysis, and interpretation of the data; or in the preparation or approval of the manuscript.

Competing interests: All authors have completed the ICMJE uniform disclosure form at www.icmje.org/coi_disclosure.pdf and declare: no support from any organisation for the submitted work; no financial relationships with any organisations that might have an interest in the submitted work in the previous three years; AB and RM are a co-chairs of the steering group for the *Preventing Overdiagnosis* scientific conferences; no other relationships or activities that could appear to have influenced the submitted work.

Data sharing: No additional data available.

REFERENCES

1. Schofield PE, Butow PN, Thompson JF, et al. Psychological responses of patients receiving a diagnosis of cancer. *Ann Oncol* 2003;**14**(1):48-56.
2. Arora NK. Interacting with cancer patients: the significance of physicians' communication behavior. *Soc Sci Med* 2003;**57**(5):791-806.
3. Hoffman JR, Cooper RJ. Overdiagnosis of disease: a modern epidemic. *Arch Intern Med* 2012;**172**(15):1123-4. doi: 10.1001/archinternmed.2012.3319
4. Welch HG, Schwartz LM, Woloshin S. Over-diagnosed: making people sick in the pursuit of health. Boston: Beacon Press 2011.
5. Moynihan R, Doust J, Henry D. Preventing overdiagnosis: how to stop harming the healthy. *BMJ* 2012;**344**:e3502. doi: 10.1136/bmj.e3502
6. Heath I. Overdiagnosis: when good intentions meet vested interests--an essay by Iona Heath. *BMJ* 2013;**347**:f6361. doi: 10.1136/bmj.f6361
7. McCaffery KJ, Jansen J, Scherer LD, et al. Walking the tightrope: communicating overdiagnosis in modern healthcare. *BMJ* 2016;**352**:i348. doi: 10.1136/bmj.i348
8. Dunn BK, Srivastava S, Kramer BS. The word "cancer": how language can corrupt thought. *BMJ* 2013;**347**:f5328. doi: 10.1136/bmj.f5328
9. Esserman LJ, Thompson IM, Reid B, et al. Addressing overdiagnosis and overtreatment in cancer: a prescription for change. *Lancet Oncol* 2014;**15**(6):e234-42. doi: 10.1016/S1470-2045(13)70598-9
10. Allegra CJ, Aberle DR, Ganschow P, et al. National Institutes of Health State-of-the-Science Conference statement: Diagnosis and Management of Ductal Carcinoma In Situ September 22-24, 2009. *J Natl Cancer Inst* 2010;**102**(3):161-9. doi: 10.1093/jnci/djp485
11. Esserman LJ, Thompson IM, Jr., Reid B. Overdiagnosis and overtreatment in cancer: an opportunity for improvement. *JAMA* 2013;**310**(8):797-8. doi: 10.1001/jama.2013.108415
12. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. *Int J Qual Health Care* 2007;**19**(6):349-57. doi: 10.1093/intqhc/mzm042

13. Schünemann HJ, Oxman AD, Higgins JPT, et al. Presenting results and 'Summary of findings' tables — results without meta-analyses 2011.
14. Omer ZB, Hwang ES, Esserman LJ, et al. Impact of ductal carcinoma in situ terminology on patient treatment preferences. *JAMA Intern Med* 2013;**173**(19):1830-1. doi: 10.1001/jamainternmed.2013.8405
15. McCaffery K, Nickel B, Moynihan R, et al. How different terminology for ductal carcinoma in situ impacts women's concern and treatment preferences: a randomised comparison within a national community survey. *BMJ Open* 2015;**5**(11):e008094. doi: 10.1136/bmjopen-2015-008094
16. Nickel B, Barratt A, Hersch J, et al. How different terminology for ductal carcinoma in situ (DCIS) impacts women's concern and management preferences: A qualitative study. *Breast* 2015;**24**(5):673-9. doi: 10.1016/j.breast.2015.08.004
17. Scherer LD, Finan C, Simanek D, et al. Effect of "Pink Eye" Label on Parents' Intent to Use Antibiotics and Perceived Contagiousness. *Clin Pediatr (Phila)* 2015 doi: 10.1177/0009922815601983
18. Scherer LD, Zikmund-Fisher BJ, Fagerlin A, et al. Influence of "GERD" label on parents' decision to medicate infants. *Pediatrics* 2013;**131**(5):839-45. doi: 10.1542/peds.2012-3070
19. Copp T, McCaffery K, Azizi L, et al. Influence of the disease label 'polycystic ovary syndrome' on intention to have an ultrasound and psychosocial outcomes: a randomised online study in young women. *Hum Reprod* 2017;1-9. doi: 10.1093/humrep/dex029
20. Azam N, Harrison M. Patients' perspectives on injuries. *Emerg Med J* 2011;**28**(7):601-3. doi: 10.1136/emj.2009.082032
21. Bedell SE, Graboyes TB, Bedell E, et al. Words that harm, words that heal. *Arch Intern Med* 2004;**164**(13):1365-8. doi: 10.1001/archinte.164.13.1365
22. Cassell EJ, Leon AC, Kaufman SG. Preliminary evidence of impaired thinking in sick patients. *Ann Intern Med* 2001;**134**(12):1120-3.
23. Chadha NK, Repanos C. Patients' understanding of words used to describe lumps: a cross-sectional study. *J Laryngol Otol* 2006;**120**(2):125-8. doi: 10.1017/S0022215105004688
24. Zikmund-Fisher BJ, Fagerlin A, Keeton K, et al. Does labeling prenatal screening test results as negative or positive affect a woman's responses? *Am J Obstet Gynecol* 2007;**197**(5):528 e1-6. doi: 10.1016/j.ajog.2007.03.076
25. Scherer AM, Scherer LD, Fagerlin A. Getting ahead of illness: using metaphors to influence medical decision making. *Med Decis Making* 2015;**35**(1):37-45. doi: 10.1177/0272989X14522547
26. Rot I, Ogah I, Wassersug RJ. The language of prostate cancer treatments and implications for informed decision making by patients. *Eur J Cancer Care (Engl)* 2012;**21**(6):766-75. doi: 10.1111/j.1365-2354.2012.01359.x
27. Pieterse AH, Jager NA, Smets EM, et al. Lay understanding of common medical terminology in oncology. *Psychooncology* 2013;**22**(5):1186-91. doi: 10.1002/pon.3096
28. Gavaruzzi T, Lotto L, Rumiati R, et al. What makes a tumor diagnosis a call to action? On the preference for action versus inaction. *Med Decis Making* 2011;**31**(2):237-44. doi: 10.1177/0272989X10377116
29. Erueti C, Glasziou P, Mar CD, et al. Do you think it's a disease? a survey of medical students. *BMC Med Educ* 2012;**12**:19. doi: 10.1186/1472-6920-12-19
30. Pickering TG. Now we are sick: labeling and hypertension. *J Clin Hypertens (Greenwich)* 2006;**8**(1):57-60.
31. Hoyt CL, Burnette JL, Auster-Gussman L. "Obesity is a disease": examining the self-regulatory impact of this public-health message. *Psychol Sci* 2014;**25**(4):997-1002. doi: 10.1177/0956797613516981

32. Haynes RB, Sackett DL, Taylor DW, et al. Increased absenteeism from work after detection and labeling of hypertensive patients. *N Engl J Med* 1978;**299**(14):741-4. doi: 10.1056/NEJM197810052991403

33. Rostrup M, Kjeldsen SE, Eide IK. Awareness of hypertension increases blood pressure and sympathetic responses to cold pressor test. *Am J Hypertens* 1990;**3**(12 Pt 1):912-7.

34. Nikiforov YE, Seethala RR, Tallini G, et al. Nomenclature Revision for Encapsulated Follicular Variant of Papillary Thyroid Carcinoma: A Paradigm Shift to Reduce Overtreatment of Indolent Tumors. *JAMA Oncol* 2016 doi: 10.1001/jamaoncol.2016.0386

35. Patel KN. Noninvasive Encapsulated Follicular Variant of Papillary Thyroid "Cancer" (or Not): Time for a Name Change. *JAMA Oncol* 2016 doi: 10.1001/jamaoncol.2016.0714

36. Independent UK Panel on Breast Cancer Screening. The benefits and harms of breast cancer screening: an independent review. *Lancet* 2012;**380**:1778-86.

37. Moyer VA, on behalf of the U.S. Preventive Services Task Force. Screening for prostate cancer: U.S. Preventive Services Task Force recommendation statement. *Ann Intern Med* 2012;**157**(2):120-34. doi: 10.7326/0003-4819-157-2-201207170-00459

38. Brito JP, Morris JC, Montori VM. Thyroid cancer: zealous imaging has increased detection and treatment of low risk tumours. *BMJ* 2013;**347**:f4706. doi: 10.1136/bmj.f4706

Table 1. Criteria for assessment of study quality for quantitative studies

	Information Extracted
Study design	Independent-sample or paired-sample design If independent: whether groups were randomised?
Study setting	Community sample, clinics, hospital, other
Selection bias	Independent sample: Were participants allocated to each group randomly? Were samples similar in terms of important characteristics? How was randomisation done? Was randomisation done with concealment allocation? Was the study sample described? Paired-sample: Were terms randomised? How was randomisation done? Was randomisation done with concealment allocation? Was the study sample described?
Performance bias	Were participants kept blind to the study hypothesis?
Attrition bias	Was attrition or exclusions (lost to follow-up) reported?
Reporting bias	Examination of selective reporting. Did they study have a protocol?
Measurement bias	Exposure variable – describe the term used Outcome measures – how were outcomes measured? Were they validated or referenced? Confounders – reporting of additional measures and/or demographics
Analysis	Were appropriate statistical tests used to analyse data and report results?

Table 2. Characteristics of included studies

Study	Year	Country	Disease Focus	Study Design	Study Sample & Setting	Study Aims	Terms Manipulated	Primary Outcomes Measured	Other Outcomes Measured
Copp et al.	2017	Australia	PCOS	Randomised experimental design (2x2x(2) factorial design)	181 female university students	To test the impact of the PCOS disease label on intention to undergo an ultrasound and psychosocial outcomes	'polycystic ovary syndrome' vs 'hormonal imbalance'	Intention to have an ultrasound	Negative affect, self-esteem, perceived severity of condition, credibility of GP, and interest in a second opinion
McCaffery et al.	2015	Australia	DCIS	Randomised experimental design (cross-over design)	269 healthy women from a community sample	To examine whether the use of terminology including the term cancer to describe DCIS increased hypothetical level of concern and treatment preferences	'abnormal cells' vs 'pre-invasive breast cancer cells'	Treatment preferences: treatment vs watchful waiting, measured on a 5-point Likert scale (definitely prefer treatment, probably prefer treatment, prefer to do nothing, probably prefer watchful waiting, definitely prefer watchful waiting)	Level of concern measured on a 5-point Likert scale (Extremely concerned-Not concerned at all)
Nickel et al.*	2015	Australia	DCIS	Semi-structured qualitative interviews	26 women from a community sample	To understand how different proposed terminologies for DCIS affect women's perceived concern and management preferences	'abnormal cells' vs 'pre-invasive breast cancer cells' 'ductal carcinoma in situ' 'ductal intraepithelial neoplasia' 'indolent lesions of epithelial origin'	Women's qualitative responses to terminologies with and without the cancer term on level of concern and management preferences	
Querner et al.	2013	USA	DCIS	Randomised experimental design (cross-over design)	394 healthy women with no history of breast cancer from a hospital patient registry	To examine how women respond to terminology for DCIS without the cancer term	'non-invasive breast cancer', 'breast lesion', 'abnormal cells'	Treatment preferences (choice between surgery, medication, active surveillance)	
Scherer et al.	2013	USA	GERD	Randomised experimental design (2x2 factorial design)	175 parents aged 18 years presenting at a primary care pediatric clinic	To determine if the disease label GERD influences parents' perceived need to medicate an infant	'GERD' vs no label 'this problem'	Parents' interest in giving their infant medication (3 measures using 6-point Likert scale: No, definitely not-Yes, definitely)	Perception of illness severity (3 measures using 5-point Likert scale: worry = Not at all worried-Very worried, serious = Not at all serious-Very serious, sick = Strongly disagree-Agree)

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49

									Appreciation of medication offered (1 measure using 5-point Likert scale: No, definitely not-Yes, definitely)
Scherer et al.	2015	USA	Conjunctivitis	Randomised experimental design (2x2 factorial design)	159 parents aged 18 years presenting at a primary care pediatric clinic	To determine whether the 'pink eye' term would influence parents' beliefs about the condition and their interest in using antibiotics	'pink eye' vs 'eye infection'	Parents' decision to give their child antibiotics measured on a 6-point Likert scale (No, definitely not-Yes, definitely)	Targeted beliefs about contagiousness measured on a 5-point Likert scale (Not at all contagious-Very contagious) Parents ability to send their child to day care measured on a 5-point Likert scale (Not at all likely-very likely)
Azam et al.	2010	UK	Bony fracture	Cross sectional survey	100 adult patients presenting at an emergency department	To assess the way different terms used to describe a fracture affect the understanding a patients has of it including the perceived severity of the injury and how the patient expects to be treated	'a crack in the bone' 'a broken bone' 'a fracture' 'a hairline fracture' 'a greenstick fracture'	Expected treatment (choice between heals on own, sling, cast, operation)	Perceived severity on a 10-point scale (1 being minimally problematic to 10 being a very serious injury)

*=included qualitative study

Table 3. Risk of bias summary for quantitative studies

Study	Study design	Study setting	Selection bias	Performance bias	Attrition bias	Reporting bias	Measurement Bias	Analysis
Copp, 2017	+	+	+	?	+	+	+	+
McCaffery, 2015	+	+	+	?	+	+	+	+
Omer, 2013	+	?	+	?	?	+	+	+
Scherer, 2013	+	+	+	?	?	+	+	+
Scherer, 2015	+	+	+	?	?	+	+	+
Azam, 2010	-	+	-	?	?	-	-	-

Table 4. Key findings from individual studies

Study	Primary Outcome	Primary Outcome by Terminology		Primary Outcome Statistical Significance*	Other Outcomes
		Medicalised term	Non-medicalised term		
Copp, 2017 (n=181)	Intention to have an ultrasound	Polycystic ovary syndrome (n=90) mean=6.62 (on a 10-point Likert scale where 1=Definitely will not to 10=Definitely will)	Hormonal imbalance (n=91) mean=5.76 (on a 10-point Likert scale where 1=Definitely will not to 10=Definitely will)	Primary outcome significant (p=0.033). Women who received 'PCOS' term in the scenario had significantly higher intentions to have an ultrasound than those who received the 'hormonal imbalance' term F(1,176)=4.63, p=0.033. After women received information on overdiagnosis both intention and perceived severity decreased, regardless of the terminology of the condition (both p<0.001).	Self-esteem: Women's self-esteem was significantly lower for those given the term 'PCOS' than those given the term 'hormonal imbalance' F(1,176)=4.74, p=0.031. Perception of severity: Women who were given the term 'PCOS' had significantly higher perceived severity about the condition than those given the term 'hormonal imbalance' F(1,176)=5.64, p=0.019. Negative affect, credibility of doctor and interest in a second opinion: No difference between terms (all p>0.05).
McCaffery, 2015 (n=269)	Treatment preference	Pre-invasive breast cancer cells (n=128) 40% (51) prefer treatment 60% (77) prefer watchful waiting Change in terminology 41% (52) prefer treatment 59% (76) prefer watchful waiting	Abnormal cells (n=141) 33% (47) prefer treatment 67% (94) prefer watchful waiting Change in terminology 45% (63) prefer treatment 55% (78) prefer watchful waiting	Primary outcomes not significant however change in terminology shows a significant difference. There were no significant difference in treatment between arm A (women who were given the term 'abnormal cells' first and then were given the term 'pre-invasive cancer cells') and arm B (women who were given the term 'pre-invasive cancer cells' first and then were given the term 'abnormal cells') with 33% and 41% of women respectively favouring treatment, p=0.23. In arm A, 18% of women changed their preference to treatment while only 6% changed to watchful waiting (p=0.008). No significant treatment preferences were observed in arm B (9% vs 8%, p>0.99).	Level of concern: There was no significant difference between arms with 49% and 44% of women indicating they would be extremely concerned p=0.600. However, when the alternative term was used women in arm A ('abnormal cells' terminology first and then 'pre-invasive cancer cells' terminology) were significantly more likely to report increased concern than women in arm B ('pre-invasive cancer cells' terminology first and then 'abnormal cells' terminology) 67% vs 52%, p=0.001.
Her, 2013	Treatment preference	Cancer 47% (186) surgery	Lesion 34% (136) surgery	Primary outcome significant (p<.001)	

1
2
3
4

5 6 7 8 9 10 11		20% (79) medication 33% (129) active surveillance	18% (70) medication 48% (188) active surveillance Abnormal cells 31% (124) surgery 21% (82) medication 48% (188) active surveillance	When DCIS was described as a 'non-invasive cancer' 53% (208 of 394) of participants preferred non-surgical options, whereas 66% (258 of 394) chose non-surgical options when it was described as 'breast lesion' and 69% (270 of 394) chose non-surgical options when it was described as 'abnormal cells' (p<.001).	
12 13 14 15 16 17 18 19 20 21 22	Parents' interest in medicating their infant (3 items pooled and mean reported)	GERD term (n=87) mean=2.51 (on a 6-point Likert scale where 0=No, definitely not to 5=Yes, definitely)	No label (n=88) mean = 2.04 (on a 6-point Likert scale where 0=No, definitely not to 5=Yes, definitely)	Primary outcome significant (p<.01) Parents who received the 'GERD' term in the scenario were more interested in medication than parents who did not receive that term, F(1,165)=6.95, p<.01. Parents not given the 'GERD' term were interested in medication only when they were not given information about medication effectiveness therefore they were allowed to assume that the medications are effective F(1,165)=4.52, P<.05.	Perception of illness severity: Findings were not influenced by the presence or absence of the 'GERD' term (all p>.12). Appreciation of medication offer: Parents were least appreciative of medication when they were told that there medication was ineffective and were not given the 'GERD' term F(1,165)=7.16, p<.01
23 24 25 26 27 28 29 30 31 32 33 34 35	Parents' decision to give their child antibiotics	Pink-eye (n=82) mean=2.43 (on a 6-point Likert scale where 0=No, definitely not to 5=Yes, definitely)	Eye infection (n=76) mean=2.32 (on a 6-point Likert scale where 0=No, definitely not to 5=Yes, definitely)	Primary outcome not significant however when adjusted for information of medication effectiveness outcomes are significant. Parents were less willing to give their child antibiotics when symptoms were referred to as an 'eye infection' information about antibiotic ineffectiveness significantly reduced F(1, 62)=14.67, p<.001. By contrast when parents were told that the symptoms were 'pink eye' interest in antibiotics was not reduced by information about antibiotic ineffectiveness F(1,74)=0.93 p=.33.	Beliefs about contagiousness: Parents who received the 'pink-eye' term thought that the symptoms were significantly more contagious than parents who received the 'eye-infection' term, F(1, 137)=11.21, p=.001. Feelings towards sending child to day care: Parents who received the 'pink-eye' term were less likely to believe that their child would be allowed to go to child care than parents who received the 'eye-infection' term, F(1, 141)=9.70, p=.002.
36 37 38 39 40	Expected treatment	Broken bone 3% heals on own 39% sling 56% cast 2% operation	Crack in the bone 24% heals on own 57% sling 13% cast 6% operation	t-tests which demonstrate significance: Crack vs break=p<0.0001 Crack vs hairline fracture=p<0.0001 Crack vs greenstick fracture=p<0.0001 Break vs hairline fracture=p=0.0001	Patients perspectives of severity: (mean/median score, no SDs given): 3.28/3 for 'a crack in the bone' 3.58/3 for 'a hairline fracture' 4.95/5 for 'a fracture'

41
42
43
44
45
46
47
48
49

1
2
3
4

5					
6		Greenstick fracture	Hairline fracture	Break vs greenstick fracture= $p<0.0001$	5.28/5 for 'a greenstick fracture'
7		47% heals on own	45% heals on own	Fracture vs hairline fracture= $p<0.0001$	6.64/7 for 'a broken bone'
8		25% sling	29% sling	Fracture vs greenstick fracture= $p<0.0001$	
9		22% cast	26% cast		
10		6% operation	0% operation		
11					
12		Fracture			
13		14% heals on own			
14		44% sling			
15		36% cast			
16		6% operation			

17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49

*as reported by original study authors
**=1 case from total sample missing

For peer review only

Table 5. Summary of preferences for more invasive management option by type of terminology*

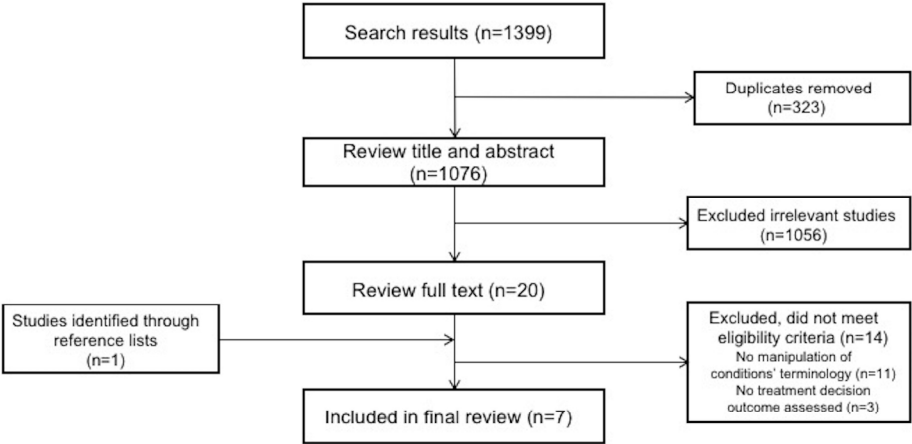
Study	More medicalised or more precise term/s (%)	Less medicalised or less precise term/s (%)	Difference (%)	p-value
Copp, 2017	70	53	17	>0.05§
McCaffery, 2015	40	33	7	0.23
Omer, 2013	47	32.5	14.5	<0.001
Scherer, 2013	74	67	7	>0.1§
Scherer, 2015	60	58	8	>0.1§
Azam, 2010	39 (4 operation, 35 cast)†	19 (6 operation, 13 cast)‡	20	<0.025§

*=combined data where applicable and mean percentages reported, see Appendix 4 for explicit justification of categorisation of terminology
†=broken bone, fracture, greenstick fracture, hairline fracture; ‡=crack in the bone
§=calculated significance using raw (Copp and Scherer) and published (Azam) data based on our classification of which terms were more medicalised

Figure 1. Flow diagram of study selection

For peer review only

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60



254x190mm (300 x 300 DPI)

Appendix 1. Search strategy

OVID Medline	OVID Pre-Medline	Embase	PsycINFO	CINAHL	PubMed
1. exp Terminology as Topic/ 2. terminolog*.tw. 3. medical term*.tw. 4. disease label*.tw. 5. (disease adj3 label*).tw. 6. medical label*.tw. 7. 1 or 2 or 3 or 4 or 5 or 6 8. Patient Preference/ 9. Decision Making/ 10. patient decision making.tw. 11. (patient adj3 decision*).tw. 12. treatment decision making*.tw. 13. (treat* adj3 decision*).tw. 14. treatment pref*.tw. 15. (treat* adj3 pref*).tw. 16. management pref*.tw. 17. (manage* adj3 pref*).tw. 18. overdiagnosis.kw. 19. 20. overtreatment.kw. 21. 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15 or 16 or 17 or 18 or 19 22. 7 and 20	1. terminolog*.tw. 2. medical term*.tw. 3. disease label*.tw. 4. (disease adj3 label*).tw. 5. medical label*.tw. 6. 1 or 2 or 3 or 4 or 5 7. patient preference/ 8. patient decision making.tw. 9. treatment decision making*.tw. 10. (treat* adj3 decision*).tw. 11. treatment pref*.tw. 12. (treat* adj3 pref*).tw. 13. management pref*.tw. 14. (manage* adj3 pref*).tw. 15. overdiagnosis.kw. 16. overtreatment.kw. 17. 7 or 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15 or 16 18. 6 and 17	1. nomenclature/ 2. terminolog*.tw. 3. medical term*.tw. 4. disease label*.tw. 5. (disease adj3 label*).tw. 6. medical label*.tw. 7. 1 or 2 or 3 or 4 or 5 or 6 8. patient preference/ 9. patient decision making/ 10. treatment decision making*.tw. 11. (treat* adj3 decision*).tw. 12. treatment pref*.tw. 13. (treat* adj3 pref*).tw. 14. management pref*.tw. 15. (manage* adj3 pref*).tw. 16. 17. overtreatment.kw. 18. 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15 or 16 or 17 19. 7 and 18	1. terminology/ 2. terminolog*.tw. 3. medical term*.tw. 4. disease label*.tw. 5. (disease adj3 label*).tw. 6. medical label*.tw. 7. 1 or 2 or 3 or 4 or 5 or 6 8. decision making/ or management decision making/ 9. patient decision making.tw. 10. treatment decision making*.tw. 11. (treat* adj3 decision*).tw. 12. treatment pref*.tw. 13. (treat* adj3 pref*).tw. 14. management pref*.tw. 15. (manage* adj3 pref*).tw. 16. overdiagnosis.kw. 17. 18. overtreatment.kw. 19. 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15 or 16 or 17 20. 7 and 18	S1 TI terminolog* or AB terminolog* S2 TI medical term* or AB medical term* S3 TI disease label* or AB disease label* S4 TI disease N3 label* or AB disease N3 label* S5 TI medical label* or AB medical label* S6 S1 OR S2 OR S3 OR S4 OR S5 S7 (MH "Decision Making") OR (MH "Decision Making, Patient") S8 TI treatment decision making or AB treatment decision making S9 TI treatment N3 decision* or AB treatment N3 decision* S10 TI treatment pref* or AB treatment pref* S11 TI treatment N3 pref* or AB treatment N3 pref* S12 TI management pref* or AB management pref* S13 TI manage* N3 pref* or AB manage* N3 pref* S14 "overdiagnosis" S15 "overtreatment" S16 S7 OR S8 OR S9 OR S10 OR S11 OR S12 OR S13 OR S14 OR S15 S17 S6 AND S16	((((terminolog*[Title/Abstract]) OR medical term*[Title/Abstract]) OR medical label*[Title/Abstract]) OR disease label*)) AND ((((treatment decision making[Title/Abstract]) OR treatment decision*[Title/Abstract]) OR treatment pref*[Title/Abstract]) OR management pref*[Title/Abstract]) OR overdiagnosis) OR overtreatment)

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Appendix 2. Eligibility criteria (with exclusion criteria)

Types of studies:	Empirical studies that considered management decision making as an anticipated or experienced outcome of being given or told one terminology vs another for the same condition including surveys, questionnaires and interviews or focus groups <i>Exclusion: review papers, editorials, commentary/discussion papers</i>
Types of participants/population:	Adults (patients, community members) aged 18 years and above including adults making surrogate decisions or decisions on behalf of their children <i>Exclusion: patients or community members less than 18 years of age making a decision for themselves</i>
Types of settings:	Any type of medical or community setting including hypothetical scenarios given to community members
Study factor (intervention):	Different term given for the same condition
Outcome factor:	Treatment or management preferences (e.g. treatment 1 vs treatment 2 vs treatment 3/no treatment)

Appendix 3. Nickel et al. qualitative study appraisal*

Were the researcher characteristics described?	Partly/moderate quality – interviews were conducted by two public health researchers with experience in qualitative research methods, however credentials, gender and characteristics of the interviewer were not reported
Was the methodological orientation to the study described?	Yes/ high quality – rationale informed by previous published DCIS terminology research (Esserman 2013, Omer 2013)
Was the target population clearly defined?	Yes/high quality – community sample of 26 Australian women aged 25 years and above who were recruited from a national community survey. Table 1 includes participant's characteristics (age, education, experience with cancer screening, previous breast cancer diagnosis, immediate family history of cancer, employment status, worked as a health professional, survey management preference)
Was the sampling strategy clearly defined?	Yes/high quality – consenting women were recruited from a national community survey and purposively selected according to their education background (lower vs higher education), previous screening experience (previously screened vs not) and management preferences (immediate treatment vs watchful waiting as indicated by their survey responses)
Were the data collection methods clearly described?	Yes/high quality – semi-structured single telephone interview which lasted 13-40 minutes, and were audio-recorded and transcribed verbatim
Was data analysis clearly described?	Yes/high quality – Framework Analysis was used to organise the data and identify main themes that capture the diverse views expressed with two researches independently coding data using a rigorous analysis process

*criteria adapted from Tong et al.¹²

Appendix 4. Justification for management and terminology classification in each study

Study	Management	Classification	Justification	Terminology	Classification	Justification
Copp, 2017*	6-10 on Likert scale "Intention to have a screening test (ultrasound)"	Interest in medical management	Likert scale with anchored end-points (1=Definitely will not; 10=Definitely will), assumed 1-5=No intention, 6-10=Intention	Polycystic ovary syndrome	More medicalised term	The condition's full medical terminology (PCOS) was deemed to be more medicalised.
	1-5 on Likert scale "Intention to have a screening test (ultrasound)"	No interest in medical management		Hormonal imbalance	Less medicalised/less precise term	
McCaffery, 2015*	Treatment	Invasive treatment		Pre-invasive breast cancer cells	More medicalised term	Terminology which included the cancer term was deemed to be more medicalised than without.
	Watchful waiting	Non-invasive treatment		Abnormal cells	Less/non-medicalised term	
Omer, 2013*	Surgery	Invasive treatment		Cancer	More medicalised term	Terminology which included the term cancer was deemed to be more medicalised than those without.
	Medication Active surveillance	Non-invasive treatment		Lesion Abnormal cells	Less/non-medicalised term	
Scherer, 2013*	2-5 on Likert scale "Will you give your infant this medication?"	Invasive treatment	Likert scale with anchored end-points (0=No, definitely not; 5=Yes, definitely), assumed 1=No, 2=Maybe not, 3=Maybe, 4=Yes†	Gastroesophageal reflux disease (GERD)	More medicalised term	The condition's full medical terminology (GERD) was deemed to be more medicalised.
	0+1 on Likert scale "Will you give your infant this medication?"	Non-invasive treatment		"This condition" (no label)	Less/non-medicalised term	
Scherer, 2015*	2-5 on Likert scale "Will you give your child antibiotics?"	Invasive treatment	Likert scale with anchored end-points (0=No, definitely not; 5=Yes, definitely), assumed 1=No, 2=Maybe not, 3=Maybe, 4=Yes†	Pink-eye	More precise medical term	Giving the condition a specific terminology (pink-eye) rather than a generic term (eye infection) was deemed to be more medicalised.
	0+1 on Likert scale "Will you give your child antibiotics?"	Non-invasive treatment		Eye infection	Less precise term	

Azam, 2010	Operation Cast	Invasive treatment		Broken bone Fracture Greenstick fracture Hairline fracture	More medicalised term	Authors determined classification based on information on the precise terminology healthcare professionals use to describe a bony injury.
	Heals on own Sling	Non-invasive treatment		Crack in the bone	Less/non- medicalised term	

*=studies where authors were specifically guided by study design to classify terminology

†=sensitivity analysis conducted



PRISMA 2009 Checklist

Section/topic	#	Checklist item	Reported on page #
TITLE			
Title	1	Identify the report as a systematic review, meta-analysis, or both.	1
ABSTRACT			
Structured summary	2	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.	2-3
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known.	5
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).	6
METHODS			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.	6
Eligibility criteria	6	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	7 (Appx.2)
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	6-7
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	6-7 (Appx.1)
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	7
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	6-7
Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.	6-7
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.	7 (Table 1 & Appx. 3)
Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means).	8
Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., I^2 for each meta-analysis)	8 (Appx. 4)



PRISMA 2009 Checklist

Page 1 of 2

Section/topic	#	Checklist item	Reported on page #
Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).	N/A
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.	
RESULTS			
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.	8
Study characteristics	18	For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.	8 (Table 2)
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12).	8-9 (Table 3)
Results of individual studies	20	For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot.	9-14 (Table 4)
Synthesis of results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.	14 (Table 5)
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see Item 15).	N/A
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]).	
DISCUSSION			
Summary of evidence	24	Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy makers).	14-15
Limitations	25	Discuss limitations at study and outcome level (e.g., risk of bias), and at review-level (e.g., incomplete retrieval of identified research, reporting bias).	16-17
Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.	18
FUNDING			
Funding	27	Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review.	19

From: Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med 6(7): e1000097. doi:10.1371/journal.pmed1000097

For peer review only - <http://bmjopen.bmj.com/site/about/guidelines.xhtml>

For more information, visit: www.prisma-statement.org. Page 2 of 2